DOCTOR OF PHILOSOPHY

O-linked N-acetylglucosamine in differentiation and gene expression of mouse and human pluripotent stem cells

Domke, Tanja Carolina Elisabeth

Award date:
2016

O-linked N-acetylglucosamine in differentiation and gene expression of mouse and human pluripotent stem cells

Tanja Carolina Elisabeth Domke

September 2016

Supervisor: Dr Marios P. Stavridis

A thesis submitted in fulfilment of the requirement for the degree of

Doctor of Philosophy

University of Dundee

School of Life Sciences

Contents

List of Figures

Abbreviations

5-caC	5-carboxymethylcytosine
5-fC	5-formylcytosine
5-hmC	5-hydroxymethylcytosine
5-mC	5-methylcytosine
ALK	activin receptor-like kinase
APC	adenomatous polyposis coli
BCV	biological coefficient of variation
bFGF	basic fibroblast growth factor
BMP	bone morphogenic protein
cDNA	complementary DNA
CTD	C-terminal domain
DAPI	4',6-diamidino-2-phenylindole
DGE	differential gene expression
DMEM	Dulbecco's minimal essential medium
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
EBs	embryoid bodies
EDTA	Ethylenediaminetetraacetic acid

EGFP . . . . . enhanced green fluorescent protein

ELISA . . . . . enzyme-linked immunosorbent assay

EpiSCs . . . . . epiblast stem cells

ES cells . . . . . embryonic stem cells

FCS . . . . . fetal calf serum

FDR . . . . . false discovery rate

Fz . . . . . Frizzled

GDF . . . . . growth differentiation factors

GMEM . . . . . Glasgow Minimal Essential Medium

GNS . . . . . GlcNAcstatin

GO . . . . . Gene ontology

GSK-3 . . . . . glycogen synthase kinase 3

HAT . . . . . histone acetyltransferase

HDAC . . . . . histone deacetylase

HGF . . . . . hepatocyte growth factor

HPLC-MS . . . high-performance liquid chromatography-mass spectrometry

ICM . . . . . inner cell mass

iPS cells . . . . . induced pluripotent stem cells

JAK/STAT . . Janus kinase/signal transducers and activators of transcription

KOSR . . . . . Knockout serum replacement

LIF . . . . . leukemia inhibitory factor

LIFR . . . . . leukemia inhibitory factor receptor

logFC . . . . . log fold change

LRP5/6 . . . . . LDL receptor related protein 5/6

MAPK . . . . . mitogen activated protein kinase

MEFs . . . . . mouse embryonic fibroblasts

MERVL . . . . . Mouse endogenous retrovirus with a leucine tRNA primer binding site

mRNA . . . . . messenger RNA

O-GlcNAc . . . O-linked N-acetylglucosamine

PBS . . . . . phosphate buffered saline

PcG . . . . . polycomb group

PCR . . . . . polymerase chain reaction

PI . . . . . propidium iodide

PolII . . . . . RNA polymerase II

PORCN . . . . . Protein-serine O-palmitoleoyltransferase porcupine

PRC . . . . . polycomb repressive complex

PTMs . . . . . posttranslational modifications

PVDF . . . . . Polyvinylidene difluoride

RFP . . . . . red fluorescent protein

RNA . . . . . ribonucleic acid

RT-qPCR . . . . reverse transcriptase quantitative polymerase chain reaction

SDS . . . . . sodium dodecyl sulfate

SEM . . . . . standard error of the mean

siRNA . . . . . small inhibitory RNA

TCF/LEF . . . T-cell factor/lymphoid enhancer factor

Tet . . . . . ten-eleven translocation

TGF . . . . . transforming growth factor

TPR . . . . . tetratricopeptide

TSA . . . . . trichostatin A

UDP . . . . . uridine diphosphate

Declarations

I declare that I am the author of this thesis and the data presented within are the result of my own research. Work from others has been consulted by me and is clearly identified in the text by relevant citations. This thesis has not been previously submitted for any form of degree at a university.

Tanja C.E. Domke

I confirm that the conditions of the Ordinance and Regulations relevant for submission of this thesis have been fulfilled.

Dr Marios P. Stavridis

Summary

Peptide posttranslational modifications have been shown to regulate multiple aspects of cell signalling, thereby influencing cellular functions. The addition of O-linked N-acetylglucosamine to serine or threonine residues (O-GlcNAcylation) of proteins has only recently been characterized and its overall role in cell signalling remains elusive to date. Recent studies suggest an essential role of O-GlcNAcylation on the viability and pluripotency of mouse and human embryonic stem (ES) cells. Here we show that increased levels of O-GlcNAcylation in response to specific inhibition of O-GlcNAc hydrolase (Oga) hinder mouse ES cell differentiation. In addition to these findings, I could also demonstrate that increased O-GlcNAcylation leads to expression of a gene set normally epigenetically repressed in mouse ES cells and associated with a subpopulation resembling cells in the 2-cell-stage embryo. I also extended our lab's investigations to human induced pluripotent stem (iPS) cells. While mesendodermal differentiation remains unaffected by high O-GlcNAc levels, neural differentiation is severely disrupted in these cells. Human iPS cells with elevated O-GlcNAcylation are unable to commit to the ectodermal lineage and fail to organize in neural tube-like structures, so-called neural rosettes. Following these observations we performed mRNA sequencing analysis on human iPS cells with high O-GlcNAc levels and found gene expression to be significantly altered. Genes affected by increased O-GlcNAcylation

include modulators of key neural developmental processes, for example components of the bone morphogenic protein signalling cascade.

1. Introduction

1.1. Pluripotent stem cells

1.1.1. Pluripotency and differentiation

Pluripotency can be described as the ability of a cell to specialize and become a cell of any cell type of the mature organism. This process of specialisation is also termed differentiation and it is a characteristic feature of embryonic stem cells (ES cells) or induced pluripotent stem cells (iPS cells). Pluripotent stem cells can differentiate into cells of the germ layers of the developing embryo as well as the germ line. During vertebrate embryonic development three germ layers can be distinguished that develop during gastrulation: ectoderm, endoderm and mesoderm (MacCord, 2013; Thomson et al., 1998). Cells of the ectoderm lineage will finally differentiate into the nervous system and the epidermis (contributing i.e. to skin and hair). Endoderm cells differentiate to form the digestive tract and inner organs, such as liver and pancreas. The skeleton and skeletal muscle, as well as the heart are derived form the mesoderm layer (Gilbert, 2013). In contrast to pluripotent cells, cell populations that are described as totipotent can contribute to all embryonic, but also to extra-embryonic tissue like the placenta. Multipotent cell types such as somatic stem cells have restricted potency and will only differentiate within their lineage.

There are two assays for assessing pluripotency. First, cells can be tested for their ability to contribute to embryonic chimeras. Pluripotent cells will be capable of differentiation into all tissues of these embryos (Robertson, 1986). The second assay is teratoma formation, an *in vivo* method which describes the formation of benign tumours in immunodeficient or syngeneic mice upon injection of cells. In this case, pluripotent cells will generate teratoma that contain fully differentiated cells of the three germ layers (Stevens and Little, 1954). Besides their ability to differentiate into the three germ layers, pluripotent stem cells exhibit unlimited self-renewal capacity and are therefore immortal (Thomson et al., 1998).

The following sections will provide background information on the two pluripotent stem cell models used in this study as well as on signalling pathways involved in maintaining pluripotency and triggering differentiation.

1.1.2. Mouse embryonic stem cells

Early mouse embryonic development is characterised by multiple sequential cell divisions of the fertilized zygote. In the first 24 hours, one cleavage occurs from which the 2-cell stage blastomere arises (Edwards and Gates, 1959). At this point, cells retain a totipotent cell potential. Importantly, at this stage the zygote's transcription is activated and the transcriptome undergoes a maternal-to-zygote switch. This process is referred to as zygotic genome activation (Bensaude et al., 1983; Schultz, 1993). The 2-cell stage of the developing embryo is also key for further development because of the reprogramming process that transforms the highly differentiated oocyte into totipotent blastomeres (Latham et al., 1992).

In the following the blastomeres undergo further divisions, mainly without cellular

growth or an increase in total cell mass (Aiken et al., 2004). First lineage decisions are made around embryonic day 2.75 (E2.75): Outer cells of the morula will become trophectoderm and give rise to extraembryonic tissue (Lawson et al., 1999), while inner cells will constitute the inner cell mass (ICM) of the early blastocyst (Boroviak and Nichols, 2014). Cells of the ICM at E3.5 represent a naïve pluripotent cell population. Those cells can be isolated and taken into monolayer culture where they represent an immortal and self-renewing population of ES cells (Evans and Kaufman, 1981; Martin, 1981).

Mouse ES cells were originally co-cultured on mouse embryonic fibroblasts. These feeder cells produce leukaemia inhibitory factor (LIF) which promotes self-renewal and aids to retain pluripotency of mouse ES cells. Today, standard culture conditions are feeder-free and mouse ES cells are maintained using LIF and fetal calf serum (FCS) as medium supplements (Smith et al., 1988; Williams et al., 1988). Instead of FCS, medium can be supplemented with bone morphogenic protein (BMP)4 (Ying et al., 2003b).

It is important to note that mouse ES cells cultured under those standard conditions (serum/LIF) represent a mixture of both naïve pluripotent cells and cells that are in a state that is more primed for differentiation (formative state) (Martin Gonzalez et al., 2016; Kalkan and Smith, 2014; Kumar et al., 2014). While cells that are in a naïve pluripotent state reflect molecular properties of the early pre-implantation blastocyst, cells in a formative state are thought to functionally correspond to a slightly later developmental stage of the blastocyst. The two subpopulations can be distinguished by marker gene expression (i.e. the naïve genes *Zfp42* and *Esrrb* are not expressed in the formative state) and they also differ slightly regarding cell potency: Naïve pluripotent stem cells contribute to blastocyst chimeras while this has not been reported to be the

case for cells in the formative state (Davidson et al., 2015; Kalkan and Smith, 2014; Hackett and Surani, 2014; Tesar et al., 2007; Guo et al., 2009). It is also possible to isolate cells in a fully primed pluripotent state which represents the post-implantation epiblast at around E5.5. Those primed pluripotent cells are called epiblast stem cells (EpiSCs). Inhibition of glycogen synthase kinase 3 (GSK-3) and mitogen activated protein kinase (MAPK) by small molecules has been shown to promote the maintenance of a homogeneous naïve pluripotent mouse ES cell population (Silva et al., 2008; Ying et al., 2008). The mouse ES cell culture condition using those two inhibitors is commonly referred to as 2i condition.

Besides the previously described naïve-formative heterogeneity of mouse ES cell cultures under serum/LIF condition, additional subpopulations have been proposed to exist. Of particular interest is a small subset of cells that have been reported to express genes characteristic for the 2-cell stage embryo (Macfarlan et al., 2012). Another study hints towards the existence of a subpopulation of cells in 2i culture conditions that express the extraembryonic endoderm marker *Hex*. This population appears to be able to form extraembryonic tissue and might therefore represent a totipotent cell type (Morgani et al., 2013).

1.1.3. Human induced pluripotent stem cells

Human iPS cells represent the second pluripotent stem cell model system used in this study. Induction of pluripotency by reprogramming is a young method that has become increasingly popular over the last decade (Takahashi and Yamanaka, 2016; Chari and Mao, 2016). Mature, fully differentiated cells are reprogrammed by introducing a combination of pluripotency related transcription factors. Initially, the

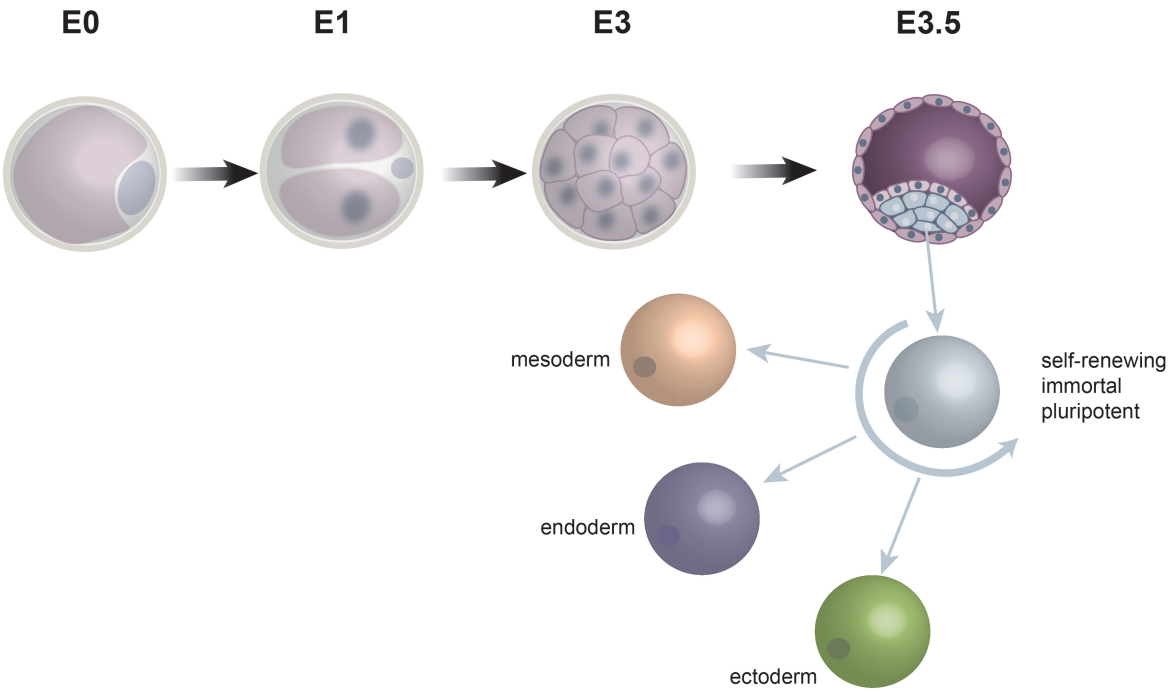

Figure (1.1) Schematic of early mouse development and mouse embryonic stem cell derivation from the blastocyst. The fertilized zygote undergoes sequential cleavage divisions. At E3.5 the embryo has developed into the pre-implantation blastocyst. At this stage, mouse ES cells can be derived from the ICM (blue) and taken into culture, where they represent a pluripotent, immortal cell population capable of self-renewal.

set of factors consisted of Oct4, Sox2, Klf4 and Myc (often referred to as OSKM), but different combinations of pluripotency transcription factors have since been used successfully to establish iPS lines (Takahashi and Yamanaka, 2006; Takahashi et al., 2007; Yu et al., 2007; Nakagawa et al., 2008; Jiang et al., 2008). These factors are simultaneously introduced into differentiated cells and will then orchestrate key reprogramming events on both transcriptional and epigenetic level. Molecular details for the mechanism underlying reprogramming are not yet understood, but cells will eventually revert to a pluripotent cell type whose functional characteristics closely resemble those of embryonic stem cells (Takahashi and Yamanaka, 2006; Takahashi et al., 2007). Respectively, iPS cells can differentiate into all mature cell types found in the adult organism. When reprogramming is complete, the ectopically expressed transcription factors are silenced and the cell will switch to expressing them endogenously (Brambrink et al., 2008; Papapetrou et al., 2009). The principle of reprogramming has been described for both mouse and human cells (Takahashi and Yamanaka, 2006; Takahashi et al., 2007; Yu et al., 2007). Importantly, generation of iPS cells is possible using different starting material, for example patient cells or cells from mutant mice, which allows investigations into multiple genetic backgrounds.

Human iPS cells are maintained in a pluripotent state by using the BMP antagonist noggin and basic fibroblast growth factor (bFGF) (Xu et al., 2005; Wang et al., 2005). These culture conditions bear the advantages of a feeder- and xeno-free environment with chemically well-defined culture conditions.

Induced pluripotent stem cells obtained by transcription factor mediated reprogramming have been demonstrated to be highly similar to ES cells. The two cell types appear to be indistinguishable in terms of their morphology, marker gene expression, global gene expression pattern and cell potency tested by teratoma formation assays

(Takahashi and Yamanaka, 2006; Takahashi et al., 2007). Nonetheless, the question whether ES cells and iPS cells are comparable model systems has been conversely debated. While several studies argue that iPS and ES cells exhibit identical properties and can not be reliably distinguished (Guenther et al., 2010; Wernig et al., 2007; Boland et al., 2009; Kang et al., 2009; Zhao et al., 2009), there are numerous reports of discrepancies between both cell types. Doi et al. (2009); Deng et al. (2009); Lister et al. (2011) and Kim et al. (2009) describe differences in deoxyribonucleic acid (DNA) methylation patterns. Furthermore, copy number variations (Hussein et al., 2011) and divergence in global gene expression profiling (Chin et al., 2009; Marchetto et al., 2009; Ghosh et al., 2010) have been reported. Several of these studies describe similarities of iPS cells with their donor cells (Ghosh et al., 2010; Marchetto et al., 2009; Doi et al., 2009).

1.1.4. Signalling pathways relevant for pluripotency

Maintenance of pluripotency and the process of differentiation are both orchestrated by multiple signalling cascades. In the following sections selected pathways will be introduced regarding their role in pluripotent stem cells. The focus will be on signalling pathways that are particularly relevant for maintenance of the undifferentiated state and that can be manipulated in order to trigger directed differentiation of pluripotent cells.

Wnt pathway

The Wnt signalling pathway is highly conserved and known to play a key role in embryonic development, cell signalling and cancer. The central component of this pathway

is β-catenin, a cytoplasmic protein that is ubiquitilated and degraded upon absence of Wnt ligands by a complex of adenomatous polyposis coli (APC), axin and GSK-3. Activation of Wnt is mediated by binding of the Wnt ligand to its receptors Frizzled (Fz) and LDL receptor related protein 5/6 (LRP5/6) which leads to GSK-3 inactivation and disruption of the β-catenin destruction complex. β-catenin will consequently accumulate and travel to the nucleus were it binds to the T-cell factor/lymphoid enhancer factor (TCF/LEF) transcription factors and initiates target gene expression (reviewed by Holland et al. (2013)).

In ES cells the Wnt pathway has been shown to promote pluripotency. In 2002, Aubert et al. (2002) demonstrated that Wnt activity attenuates neural differentiation by inhibiting GSK-3 using lithium chloride. Consistent results were obtained by Sato et al. (2004), who reported that active Wnt signalling is sufficient to sustain pluripotency in both human and mouse ES cells.

LIF-mediated JAK/STAT signalling

Janus kinase/signal transducers and activators of transcription (JAK/STAT) signalling activity can be initiated by binding of the cytokine LIF to its receptor constituted by leukaemia inhibitory factor receptor (LIFR) and gp130. Consequently, several intracellular signalling cascades are activated including the JAK/STAT pathway. LIF binding results in phosphorylation of the receptor heterodimer, followed by recruitment of STAT1 and STAT3. Both STAT proteins undergo JAK-mediated phosphorylation activation and thereupon dimerize and translocate to the nucleus where they act as transcriptional activators (Okita and Yamanaka, 2006).

Concerning mouse ES cell culture conditions, the addition of recombinant LIF to culture media containing serum or BMP4 has been shown to inhibit differentiation,

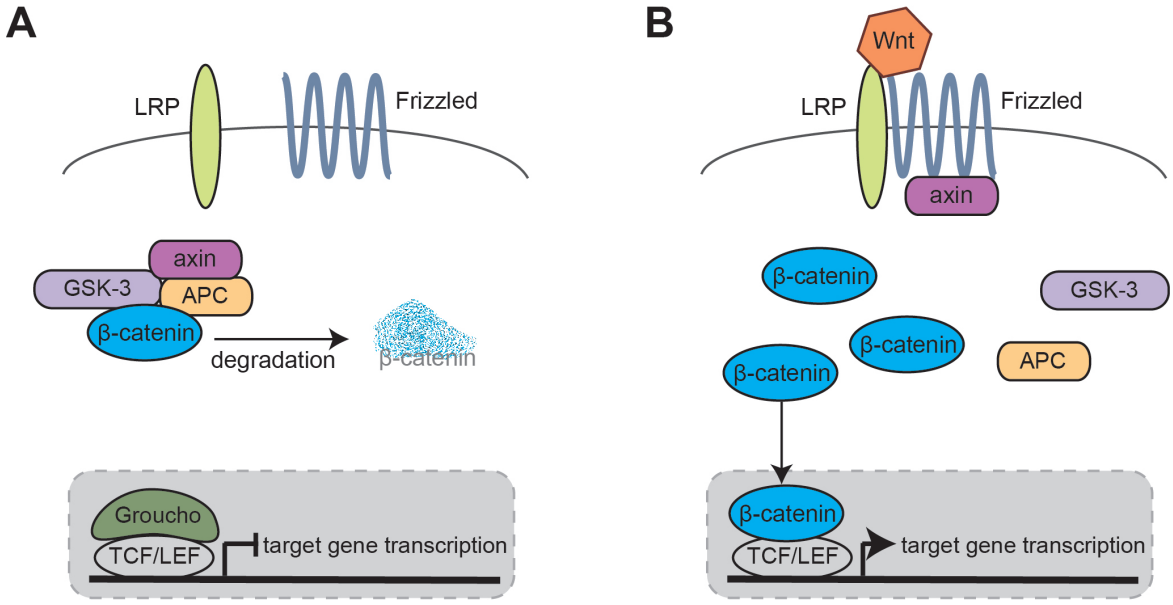

Figure (1.2) The canonical Wnt pathway. A: In absence of Wnt ligand β-catenin is bound by a complex containing APC, axin and GSK-3. GSK-3 phosphorylates β-catenin which is consequently degraded. **B:** Upon ligand binding axin is relocated to the plasma membrane. As a result the destruction complex is disrupted and GSK-3 no longer marks β-catenin for degradation. β-catenin accumulates and translocates to the nucleus where it acts as a co-activator of TCF/LEF transcription factors.

rendering the previously common feeder layers gratuitous (Smith et al., 1988; Williams et al., 1988). This effect has been reported to be due to LIF-induced STAT1/3 signalling activity, as proven by several independent studies (Boeuf et al., 1997; Niwa et al., 1998; Matsuda et al., 1999; Raz et al., 1999). These studies indicate that STAT1/3 signalling plays a key role in pluripotency maintenance, however it has been shown to not be sufficient to protect mouse ES cells from differentiation (Ying et al., 2003a). In contrast to its role in mouse ES cells, LIF-mediated STAT1/3 activity has been reported to be dispensable in human and monkey ES cells (Sumi et al., 2004; Humphrey et al., 2004). Taken together, these results indicate that while STAT1/3 activity makes a major contribution to maintaining mouse ES cells in an undifferentiated state, pluripotency maintenance in human ES is achieved by a LIF/STAT3 independent mechanism (Humphrey et al., 2004; Sumi et al., 2004).

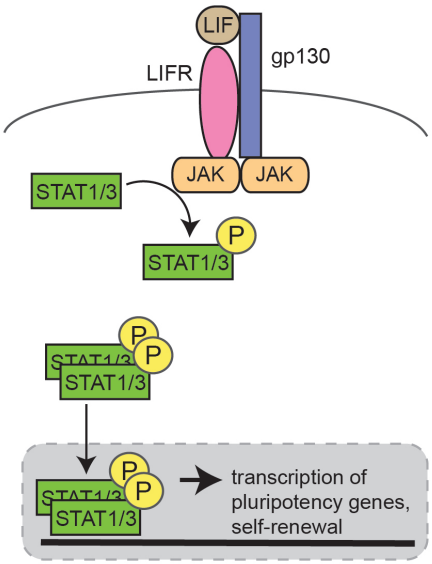

Figure (1.3) LIF-mediated STAT signalling. Binding of LIF to the heterodimer LIFR/gp130 results in receptor phosphorylation which recruits STAT1/3. JAK tyrosine kinases are bound to LIFR/gp130 and phosphorylate STAT1/3 which dimerizes and translocates to the nucleus where it acts as a transcriptional activator for pluripotency genes and thereby supports self-renewal of ES cells.

TGF-β pathway and BMP signalling

The transforming growth factor (TGF)-β superfamily has been demonstrated to influence many aspects of embryonic development as well as stem cell signalling (reviewed by Wu and Hill (2009) and Oshimori and Fuchs (2012)). Multiple ligands can activate the signalling cascade: TGF-β 1-3, Activins, Nodal, BMPs and growth differentiation factors (GDF). Receptors are constituted of two subunits which are brought together upon ligand binding and act as transmembrane serine-threonine kinases (Derynck, 1994). The cellular effectors that become phosphorylated in this process are termed regulatory Smads (R-Smads). Phosphorylated R-Smads can then form a complex with Smad4 which translocates to the nucleus where it acts as a transcription factor (Oshimori and Fuchs, 2012). Depending on the ligand that activated the pathway, different receptors and combinations of R-Smads are affected downstream: TGF-β 1-3, Activin

or Nodal binding induces phosphorylation of Smad2/3, while BMP-mediated pathway activation results in Smad1/5/8 phosphorylation (Oshimori and Fuchs, 2012). Smad6/7 are prevent phosphorylation of R-Smads and are therefore termed inhibitory Smads (I-Smads) (Nakao et al., 1997; Imamura et al., 1997).

With a view to maintaining pluripotent stem cells in an undifferentiated state, signalling through BMP4 is of particular importance. In mouse ES cells, BMP4 maintains a pluripotent state and is therefore used as a supplement of the culture medium to replace serum (Ying et al., 2003b; Qi et al., 2004). Besides Smad1/5/8 phosphorylation, BMP4 can inhibit mitogen-activated kinase pathways which has been suggested to be the underlying mechanism for its role in maintaining pluripotent cells (Qi et al., 2004). Strikingly, BMP4 signalling has been show to have an opposing effect in human ES cells: Instead of maintaining pluripotency it initiates differentiation and blockage of BMP4 signals by the antagonist noggin is necessary to preserve an undifferentiated state (Xu et al., 2002; Wang et al., 2005; James et al., 2005; Kurek et al., 2015).

Directed differentiation of pluripotent stem cells

Modulation of the above mentioned pathways can be used in order to stimulate differentiation into specific lineages and cell types. For the study at hand, directed differentiation assays into cells of ectoderm, mesoderm and endoderm lineage have been performed. There are a multitude of respective protocols and differentiation can be driven until late stages of cell maturation. Therefore, the subsequently presented protocols only represent three examples of how pathway manipulation can be used to steer pluripotent stem cell differentiation in human iPS or ES cells.

Neural differentiation has been demonstrated to be successfully driven by inhibition of TGF-β signalling through activin/nodal (Smith et al., 2008). The protocol used

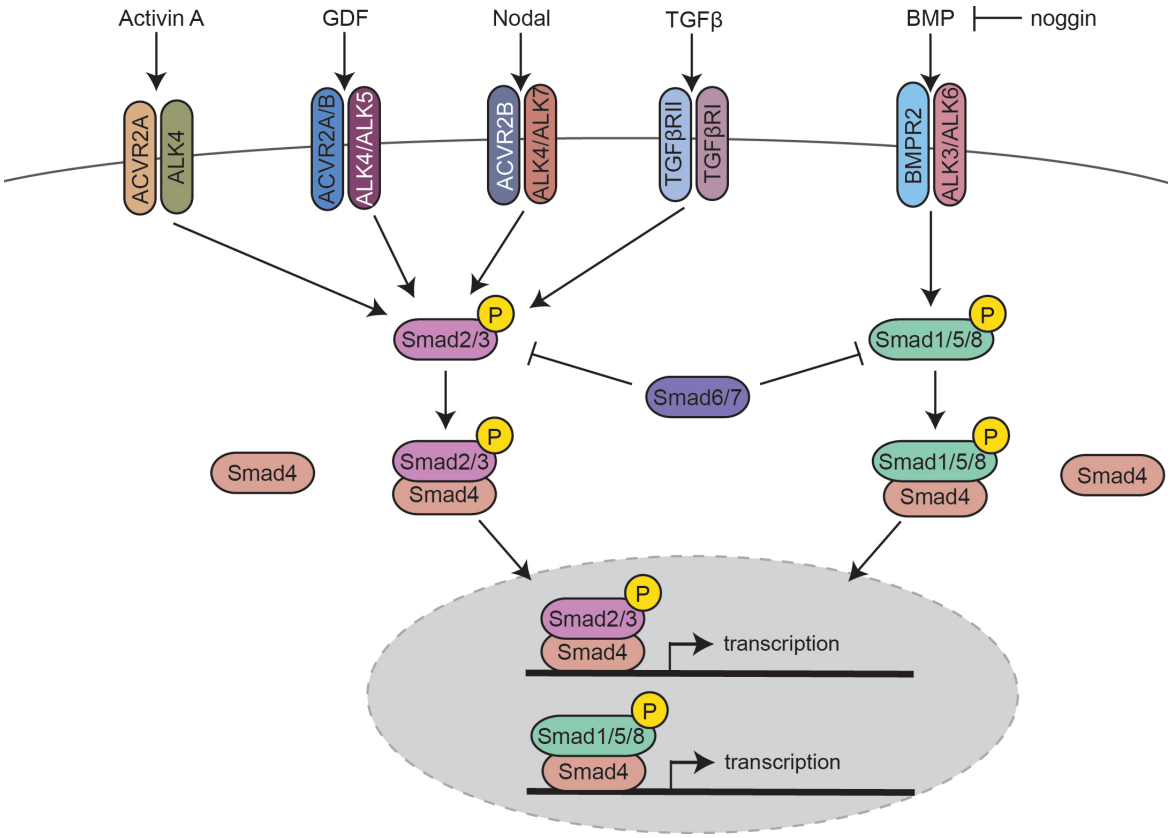

Figure (1.4) Overview of TGF-β signalling. TGF-β signalling is activated by ligand binding to respective receptor pairs which act as kinases phosphorylating R-Smads (Smad2/3 or Smad1/5/8). Phosphorylation can be blocked by inhibitory Smads (Smad6/7). After complex formation with Smad4, Smad2/3 or Smad1/5/8 enter the nucleus and drive transcription.

for neural differentiation in the study applies inhibition of two routes of the TGF-β pathway by blocking both signalling through activin/nodal and BMP (Chambers et al., 2009). The small molecule SB431542 blocks phosphorylation of the receptor subtypes activin receptor-like kinase (ALK)4, ALK5 and ALK7, which in turn results in inhibition of Smad2/3 phosphorylation (Laping et al., 2002). At the same time, the BMP antagonist LDN-193189 inhibits the type I receptors ALK2, ALK3 and ALK6 and thereby blocks phosphorlyation of Smad1/5/8 (Cuny et al., 2008). It is not entirely clear how this dual-Smad-inhibition acts synergistically to promote neural differentiation in human pluripotent stem cells. However, it has been proposed that TGF-β signalling through activin directly acts on the pluripotency network (Xu et al., 2008) and blockage of this pathway could release the cells from their pluripotent state. Parallel inhibition of BMP signalling could then hinder differentiation into trophoblast (Xu et al., 2002) or mesendoderm (D'Amour et al., 2005; Laflamme et al., 2007) and therefore steer differentiation towards a neural fate.

Differentiation into definitive endoderm can be achieved by combined modulation of Wnt and activin signalling (D'Amour et al., 2006). According to the protocol used for the present study, cells are subjected to high doses of Activin A in order to activate TGF-β signalling and to support differentiation towards mesendoderm (Medine et al., 2011). In addition, the Wnt signalling pathway is activated by supplementing the culture medium with Wnt3a. Alternatively, Wnt pathway activation can be achieved by using a small molecule inhibitor targeting GSK-3 (Kunisada et al., 2012). Further studies have demonstrated that Wnt signalling can be substituted for by adding BMP (Teo et al., 2012; Vallier et al., 2009). How exactly the cooperation of Activin signalling with Wnt or BMP acts on a molecular level to drive definitive endoderm differentiation has not been resolved to date.

In order to achieve differentiation of human pluripotent stem cells into mesoderm, modulation of the Wnt signalling cascade can be used. For this study, differentiation into cardiomyocytes was performed according to the protocol described by Burridge et al. (2014). Initially, the Wnt pathway is stimulated by blocking GSK-3 using a small molecule inhibitor. After a defined period (i.e. two days), the Wnt pathway is blocked by adding the Protein-serine O-palmitoleoyltransferase porcupine (PORCN) inhibitor Wnt-C59 (Proffitt et al., 2013) for another two days. Successful derivation of cardiomyocytes is also depending on composition of the basal medium and can be promoted by providing an extracellular matrix (Zhang et al., 2012).

Overall it is important to note that multiple factors interact during directed differentiation of pluripotent stem cells. The previously described examples illustrate the impact of signalling pathways and possibilities of manipulating these in order to favour commitment to certain lineages. However, successful differentiation is also highly depending on cell density, extracellular matrices provided and composition of the basal medium used during the assay.

1.2. O-GlcNAc: a dynamic posttranslational modification

Protein posttranslational modifications (PTMs) have been shown to regulate multiple aspects of cell signalling, thereby influencing cellular functions. A wide range of PTMs are known to date and many of them, i.e. phosphorylation, methylation or ubiquitination, have been extensively studied. O-linked N-acetylglucosamine (O-GlcNAc) describes the addition of a single, uncharged N-acetylglucosaminyl residue to hydroxyl groups of serine or threonine by an O-glycosidic linkage. This PTM has been characterized around 30 years ago (Torres and Hart, 1984) and has been demonstrated

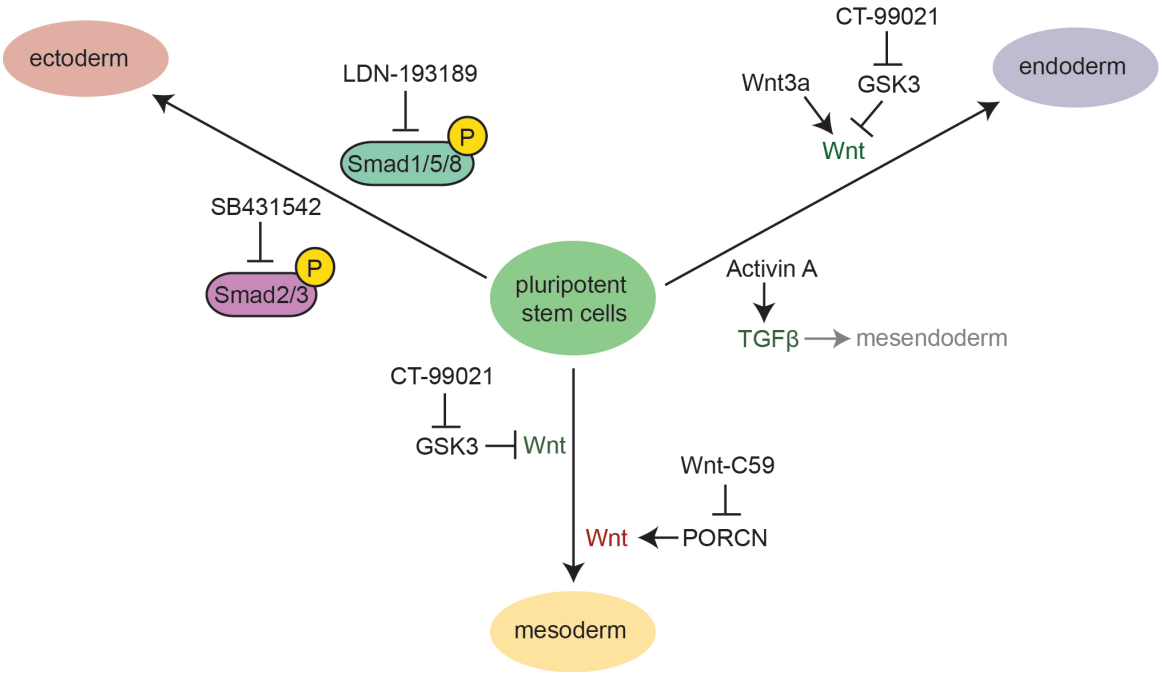

Figure (1.5) Overview of signalling modulation used for this study in order to direct pluripotent stem cell differentiation. Ectoderm (neural) differentiation can be achieved by dual inhibition of Smad phosphorylation. Differentiation into definitive endoderm can be triggered through activation of both TGF-β (through Activin A) and Wnt signalling. With the aim of differentiating into mesoderm the Wnt pathway is initially activated and subsequently blocked.

to be distinct from other forms of glycosylation in terms of kinetics, substrates and characteristics of its regulators. For example, O-GlcNAc does not occur extracellularly and is not usually elongated by additional molecules (Hart and Akimoto, 2009). Furthermore, O-GlcNAc is a dynamic and highly abundant, intracellular modification that has been found in the nucleus and cytoplasm of all eukaryotes examined so far (Kreppel et al., 1997; Lubas et al., 1997). Interestingly, there is reported interplay between O-GlcNAcylation and phosphorylation as most O-GlcNAc modified proteins are also phophoproteins. Both modifications occur on Ser/Thr residues and have been shown to occasionally compete. Respective sites are therefore often termed Yin-Yang sites (Rexach et al., 2010). In contrast to phosphorylation, which is mediated by more than 500 protein kinases (Manning et al., 2002) and about 200 phosphatases (Sacco et al., 2012), O-GlcNAc cycling is catalysed by only two enzymes. While the addition of O-GlcNAc is administered by the O-GlcNAc transferase Ogt, its removal is regulated by the hydrolase Oga (also known as Mgea5) (Gao et al., 2001; Kreppel et al., 1997; Lubas et al., 1997).

O-GlcNAc has a key function as nutrient sensor: The product of the nutrient-sensitive hexosamine biosynthetic pathway, the nucleotide monosaccharide uridine diphosphate (UDP)-GlcNAc, acts as a substrate for the O-GlcNAc transferase Ogt. Availability of UDP-GlcNAc is therefore depending on nutritional input and the abundance of the substrate can directly impact on O-GlcNAc levels (reviewed in Wells et al. (2003)). While O-GlcNAc has been implicated to play a major role in metabolism, diabetes and Alzheimer's (Dias and Hart, 2007) very little is known to date about its function in stem cell signalling and maintenance. The following sections will introduce basic regulation of O-GlcNAc and provide information on stem cell relevant molecular mechanisms that are influenced by the modification.

1.2.1. Regulating enzymes

Both enzymes that are responsible for mediating attachment and removal of the modification are highly abundant in the cell. Importantly, both have also been shown to be highly conserved throughout species (Kreppel et al., 1997). In addition, Ogt and Oga are themselves targets of O-GlcNAcylation, but also phosphorylation and most likely regulated by those modifications (Lubas and Hanover, 2000; Beausoleil et al., 2004; Khidekel et al., 2007).

Ogt: the O-GlcNAc transferase

The O-GlcNAc transferase Ogt mediates the attachment of O-GlcNAc to target proteins (Haltiwanger et al., 1990). The gene encoding the transferase is located on the X chromosome in both mouse and human and contains two potential promoter sequences that could lead to the generation of four transcripts, however, only three isoforms have been reported to date (Shafi et al., 2000; Hanover et al., 2003). The enzyme is highly abundant in most cells. Importantly, the expression of the three different isoforms varies in a tissue-dependent manner (Kreppel et al., 1997). The largest Ogt isoform (116 kDa) is the nucleo-cytoplasmic Ogt (ncOgt), followed by a mitochondria-specific 103 kDa protein (mOgt) and a small isoform with 78 kDa (sOgt) (Lubas et al., 1997; Love et al., 2003; Hanover et al., 2003). The N-terminal domain of Ogt contains multiple tetratricopeptide (TPR) repeats and the three isoforms are mainly distinguished by variation in TPR number: ncOgt contains 12 TPRs, mOgt 9 TPRs and sOgt only 2 TPRs (Hanover et al., 2003). A potential function of the Ogt N-terminus is mediating protein-protein interactions (Kreppel and Hart, 1999; Kreppel et al., 1997; Jinek et al., 2004), but it is also to some extent required for the enzymatic activity of the protein

(Lubas and Hanover, 2000). The C-terminal domain of Ogt harbours the active site of the enzyme that binds UDP-GlcNAc and catalyses the attachment to target proteins (Lazarus et al., 2011; Kreppel and Hart, 1999).

Ogt appears to be regulated on multiple levels, but full mechanistic details remain largely elusive. First of all, the protein has been shown to exhibit autoglycosylation ability and can be O-GlcNAc modified both within the catalytic domain and the TPR repeats (Kreppel et al., 1997; Lubas and Hanover, 2000; Tai et al., 2004). Ogt can also be tyrosine-phosphorylated (Kreppel et al., 1997), potentially within the TPR region (Kreppel and Hart, 1999). The role of these modifications has not yet been elucidated.

Furthermore, Ogt is regulated by nutrient input through the availability of UDP-GlcNAc (Kreppel et al., 1997; Kreppel and Hart, 1999). Increasing UDP-GlcNAc levels also increase Ogt's affinity to the substrate, while free UDP has an inhibitory effect (Haltiwanger et al., 1990; Kreppel et al., 1997; Kreppel and Hart, 1999). Moreover, Kreppel and Hart (1999) found that Ogt can form multimers (both homotrimers and heterotrimers of different isoforms). Interestingly, multimerisation does not impact on specific catalytic activity, but can alter substrate affinity.

Ogt also interacts with numerous proteins via the TPR domain. These protein-protein interactions are likely to influence the activity of Ogt, its localization as well as its ability to bind to further interaction partners and contribute to complex formation (Jinek et al., 2004; Lubas and Hanover, 2000). For example, Ogt has been demonstrated to form a complex with Sin3A which can repress transcription. In this context, it has been proposed that Sin3A acts to anchor Ogt to relevant promoters (Yang et al., 2002). Interestingly, Ogt has also been shown to form a complex with its opposite, the O-GlcNAc hydrolase Oga (Whisenhunt et al., 2006). This phenomenon is biologically highly unusual, however, its effects have not been further studied to

date.

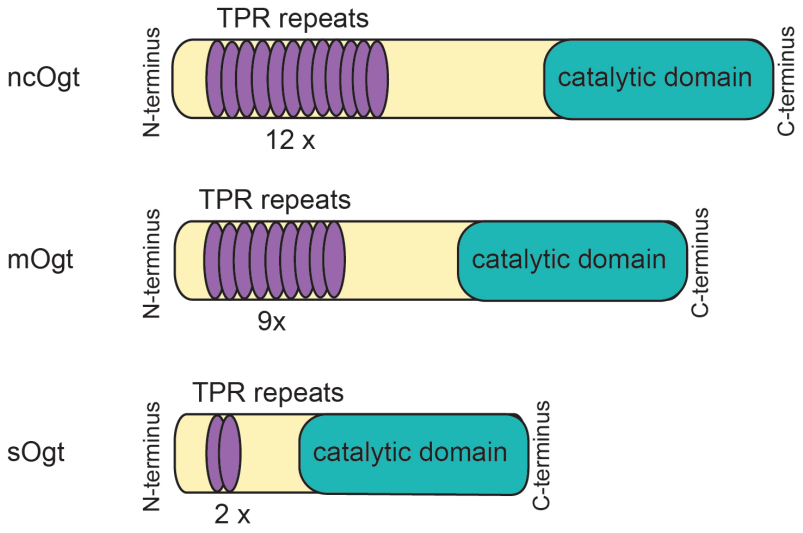

Figure (1.6) The O-GlcNAc transferase Ogt. Structure of the three Ogt isoforms: nucleo-cytoplasmic (nc), mitochondria-specific (m) and small (s) Ogt.

Oga: the O-GlcNAc hydrolase

As previously mentioned the O-GlcNAc hydrolase Oga catalyses the removal of O-GlcNAc from target proteins. The gene encoding Oga has been first characterized as meninginoma expressed antigen 5 (MGEA5) and resides on chromosome 10 (Heckel et al., 1998). An alternative splicing event can result in the expression of a smaller variant of Oga (Oga-SV) (Heckel et al., 1998). While full-length Oga is located mainly in the cytoplasm, Oga-SV has been demonstrated to primarily accumulate in the nucleus (Gao et al., 2001; Comtesse et al., 2001). The splicing variant is truncated on the C-terminus and lacks a region that has been proposed to harbour a histone acetyltransferase (HAT) domain (Schultz, 2002). While it has been reported that this HAT domain is functional (Toleman et al., 2004), there are also opposing reports (including the data presented in this study) that indicate that the domain is catalytically inactive

and represents a pseudo-HAT domain (He et al., 2014; Rao et al., 2013). The catalytic centre responsible for the hexosaminidase activity of Oga is located at the N-terminus of the enzyme (Schultz, 2002).

It remains elusive to what extent and under which mechanisms the activity of Oga can be regulated. There has been a report describing cleavage of Oga by caspase 3, however, the cleaved enzyme fragments retain O-GlcNAcase activity (Butkinaree et al., 2008). Oga has also been demonstrated to be subject to both phosphorylation (Ser 364) and O-GlcNAcylation (Ser 405), but no functional consequences have been associated with these modifications (Beausoleil et al., 2004; Khidekel et al., 2007). The activity of Oga is apparently also regulated by substrate availability as GlcNAc acts as competitive inhibitor of the hexosaminidase (Gao et al., 2001). Furthermore, Oga could be regulated by interaction and complex formation with other proteins. However, this aspect has been poorly studied and the most prominent interaction partner of Oga appears to be Ogt. Whisenhunt et al. (2006) report that O-GlcNAcase activity of Oga is reduced when in complex with Ogt in *in vitro* assays.

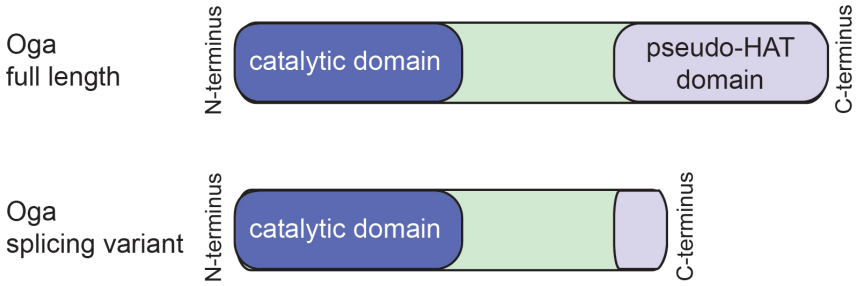

Figure (1.7) The O-GlcNAc hydrolase Oga. Structure of the two Oga variants. The full-length protein contains a pseudo-HAT domain at the C-terminus which is truncated in a natural splicing variant.

1.2.2. O-GlcNAc in pluripotent stem cells

With a view to elucidating the role of O-GlcNAc in pluripotent stem cells, its effects on the differentiation capability of these cells and their pluripotency can be proclaimed as the most crucial element in this field. Surprisingly, this has so far been a poorly studied aspect.

Nevertheless, a few studies provide evidence that O-GlcNAc modifications play a major part in differentiation. For instance, Kim et al. (2009) describe a progressive reduction of O-GlcNAc levels during cardiomyocyte differentiation of mouse ES cells. Moreover, their results indicate that an increase in O-GlcNAcylation impairs cardiogenesis. In accordance to this study, Speakman et al. (2014) confirm that O-GlcNAc levels decrease during mouse ES cell differentiation. In addition, this study indicates that Oga inhibition using the highly specific small molecule inhibitor GlcNAcstatin (GNS) C hinders differentiation of mouse ES cells into all lineages. However, early transition from a naïve state of pluripotency to a primed state appears to be unaffected by inhibition of Oga. The respective data is partly included in the thesis at hand. Only one study has so far focused on the effects of O-GlcNAc in human pluripotent stem cell differentiation: Maury et al. (2013) used both human ES and iPS cells to investigate effects of Oga inhibition using the hexosaminidase inhibitors PUGNAc and Thiamet G in undifferentiated cultures and during differentiation. Their findings indicate that pluripotency is not affected by Oga inhibitor treatments under standard culture conditions. When investigating spontaneous differentiation in embryoid bodies they report that some lineage markers are deregulated in cells treated with PUGNAc or Thiamet G. The ectoderm marker genes Pax6 and Msx1 appear to be downregulated, while the adipose-related (endoderm) markers PPAR γ and C/EBP α seem to

be upregulated at some point of differentiation as a consequence of Oga inhibition. In conclusion, it seems likely that the removal of O-GlcNAc from some proteins is required for correct ES cell differentiation. It might be speculated that the removal of O-GlcNAc from key pluripotency transcription factors plays a role in this context.

Interestingly, some of these transcription factors have previously been reported to be O-GlcNAc modified. In 1995, Chou et al. (1995) showed that c-Myc is O-GlcNAcylated at T58. However, no functional consequences of this modification have so far been addressed, although the site is also subject to phosphorylation by GSK3 and therefore appears to be a site of competition between the two modifications. In addition to this, three O-GlcNAyclation sites of Sox2 have been identified (S248, T258, S259) (Myers et al., 2011; Jang et al., 2012). Again, no functional consequences have been assigned. Of special interest is the modification of Oct4 by O-GlcNAc. Two different sites of Oct4 have been demonstrated to be O-GlcNAc modified (T228, S229) (Jang et al., 2012). Strikingly, the O-GlcNAc modification of T228 appears to have a major impact on pluripotency. By using a mutant form of Oct4 that was defective for O-GlcNAcylation, Jang et al. (2012) could demonstrate that O-GlcNAc modification of Oct4 is essential to maintain pluripotency and has a key role regarding reprogramming efficiency. To date, no reports of O-GlcNAcylation of Nanog or Klf4 have been published. However, recent unpublished data from our lab indicate that Klf4 is an O-GlcNAc target.

O-GlcNAc in stem cell relevant pathways

Wnt pathway In ES cells the Wnt pathway has been shown to maintain pluripotency. In 2002, Aubert et al. (2002) demonstrated that Wnt activity attenuates neural differentiation in mouse ES cells by inhibiting GSK-3 using lithium chloride. Consistent results were obtained by Sato et al. (2004), who reported that active Wnt

signalling is sufficient to sustain pluripotency in human and mouse ES cells. To date, the exact mechanism by which O-GlcNAcylation of Wnt signalling molecules influences ES cell signalling is still not fully understood. So far only β-catenin has been shown to be O-GlcNAc modified (Sayat et al., 2008) and the consequences for pluripotent stem cell functions are currently still under debate. While Sayat et al. (2008) report that O-GlcNAc-modification of β-catenin in human CaP cells is inversely correlated to its nuclear localization levels and therefore decreases its transcriptional activity, our more recent results clearly indicate increased Wnt signalling and elevated transcriptional activity of β-catenin upon Oga inhibition (Marios Stavridis, unpublished data). These results are consistent with a recent study in mice and various cell lines demonstrating an increase in β-catenin protein levels upon elevation of O-GlcNAc content (Olivier-Van Stichelen et al., 2012). Taken together these results might implicate a role of O-GlcNAc in pluripotency maintenance via Wnt signalling activity which could possibly be mediated by the modification of β-catenin. However, the underlying mechanism remains elusive.

LIF-mediated STAT3 signalling In 2004, Gewinner et al. (2004) reported for the first time that STAT3 is O-GlcNAcylated and that interaction of STAT family proteins with transcriptional co-activators can be influenced by their O-GlcNAcylation, implying a role of O-GlcNAc in regulation of the transcriptional activity of STATs. Several years later, Whelan et al. (2008) confirmed the O-GlcNAc modification of STAT3. Furthermore, they were able to show that Ogt activity increases in response to elevated insulin levels in adipocytes. In contradiction to these results, data from our lab indicate no changes in STAT3 activating phosphorylation upon increased O-GlcNAc levels in mouse ES cells (Speakman et al., 2014). As our study is the first to address

the impact of O-GlcNAc on STAT3 signalling in ES cells, this result could implicate that there are no direct effects of O-GlcNAc on STAT3. Nonetheless, O-GlcNAc might modify targets downstream of STAT3. As an example, it has been shown that c-Myc, a STAT3 target gene, is O-GlcNAcylated (Chou et al., 1995). Overall, the role of O-GlcNAc in ES cells STAT3 signalling is poorly understood and further work is required to clarify the effect of O-GlcNAcylation on this pathway.

TGF-β and BMP signalling Overall, very little is known about potential crosstalk between O-GlcNAcylation and TGF-β/BMP signalling. Only a single study has so far described an interaction on the levels of R-Smads: Herhaus (2014) described Ogt as a potential interaction partner of the R-Smad Smad2 in HeLa cells. Furthermore, this study indicated that Smad4 can be O-GlcNAc modified in HEK293 or HeLa cells. However, functional consequences of the modification for respective cellular signalling have not been addressed.

O-GlcNAc in transcriptional regulation and epigenetics

Recent studies on O-GlcNAcylation put emphasis on its role in epigenetic regulation. While consequences of the epigenetic functions of O-GlcNAc for ES cells remain elusive, first hints indicate a key role in transcriptional regulation as well as in repression mechanisms and are therefore likely to have an impact on ES cell self-renewal and pluripotency.

Histone code Apart from its interaction with chromatin remodelling factors, O-GlcNAc modification is also part of the histone code. By using mass spectrometry, Sakabe et al. (2010) identified sites of O-GlcNAcylation of H2A, H2B and H4. More

recently, further studies by Zhang et al. (2011) and Fong et al. (2012) indicated that H3 is also modified by O-GlcNAc. Overall, the following sites of O-GlcNAc modification of histones have been identified: H2AT101, H2BS36, H2BS112, H3S10 and H4S47 (Sakabe et al., 2010; Zhang et al., 2011; Fujiki et al., 2011; Fong et al., 2012). The consequences of O-GlcNAcylation of many of the identified sites remain elusive to date. Nonetheless, it has been shown that the O-GlcNAc sites H3S10, H2BS36 and H4S47 are also subject to phosphorylation, implying that both PTMs might compete in this context (Sakabe et al., 2010). Overall, histone modifications influence chromatin accessibility by affecting its configuration or by mediating the recruitment of enzymes (Kouzarides, 2007). In pluripotent stem cells, the chromatin configuration appears to be in an accessible state, displaying a decrease of repressive and an increase in active histone modifications in comparison to differentiated cells (Hawkins et al., 2010; Wen et al., 2009; Krejci et al., 2009). Some of the O-GlcNAc modifications on histones mentioned above are likely to have a function in pluripotent stem cell maintenance associated with respective influences on the chromatin configuration. Firstly, H3S10 is a site that has been previously implicated to mediate both repressive and active chromatin status (Zhang et al., 2011). In addition, O-GlcNAc modification of H2BS112 has been shown to facilitate this histone's monoubiquitination at K120 which has been implicated with transcriptional activation (Fujiki et al., 2011). These results indicate a possible function of O-GlcNAc in regulating transcriptional activity by modifying histones which might contribute to an open chromatin status in pluripotent stem cells.

O-GlcNAcylation of RNA polymerase II The C-terminal domain (CTD) of RNA polymerase II (PolII) is known to be modified by several kinases with major effects on its function as a transcriptional regulator, but also on complex formation with

mRNA capping enzymes, splicing and polyadenylation factors (Hirose and Manley, 2000; Comer and Hart, 2001). In 1993, Kelly et al. (1993) demonstrated that the unphosphorylated CTD of PolII is also subject to O-GlcNAcylation. Findings presented in this study also indicate that, in this case, phosphorylation and O-GlcNAcylation are mutually exclusive. Further studies suggest that different PTMs on the CTD might give rise to functionally distinct forms of PolII (Comer and Hart, 2001). However, detailed consequences on epigenetic functions are not yet fully understood. PolII is an essential regulation element in development. Subsequent to transcription initiation, PolII can pause directly downstream of transcription initiation sites and remain in a poised state. Poised PolII can be found at developmental control genes and has recently been suggested to prepare genes for rapid expression in the course of development as demonstrated by Gaertner et al. (2012). Interestingly, this study also implicates an interplay between temporally regulated poised PolII and tissue-specific Polycomb repression. These findings are in line with results obtained by Chopra et al. (2011) suggesting that poised PolII levels are regulated by polycomb group (PcG) repressors in the *Drosophila* embryo. Strikingly, the *Drosophila* homologue for O-GlcNAc transferase Ogt, *super sex combs* (sxc), is a member of the polycomb group (Sinclair et al., 2009) (see section 1.2.2). Taken together the aforementioned results imply a key role of O-GlcNAc in regulation of PolII functional states and thereby in gene expression at different developmental stages. Although not yet proven, an effect on pluripotent stem cells is probable as PolII epigenetic regulation mechanisms are known to be highly conserved.

O-GlcNAc and epigenetic repression DNA methylation is another essential epigenetic mechanism to regulate gene expression. Ten-eleven translocation (Tet) pro-

teins have recently been reported to play a crucial role in DNA methlyation. The mammalian family of Tet proteins comprises three members: Tet1, Tet2 and Tet3 (Iyer et al., 2009). Tet proteins mediate the conversion of 5-methylcytosine (5-mC) to 5-hydroxymethylcytosine (5-hmC), eventually leading to DNA demethylation (Tahiliani et al., 2009; Guo et al., 2011) and all three Tet family members have recently been reported to interact with Ogt (Deplus et al., 2013; Ito et al., 2013; Shi et al., 2013). Interestingly, Tet-mediated demethylation is vital for the recruitment of the polycomb repressive complex (PRC)2 (compare section 1.2.2) (Wu and Zhang, 2011). Several recent studies connect O-GlcNAc to both Tet protein function and pluripotent stem cell epigenetics. Vella et al. (2013) identified both Tet1 and Tet2 as stable partners of Ogt in mouse ES cells. Furthermore, the results obtained in this study indicate that Ogt and Tet1 associate at CpG-rich promoter regions and imply that Tet1 might exhibit an anchor-like function thereby mediating Ogt recruitment to chromatin. Intriguingly, further results obtained by Vella et al. (2013) also suggest the existence of a multiprotein complex consisting of Ogt, Tet1, Tet2, Sin3A and histone deacetylase (HDAC)1. As both Sin3A and HDAC1 are known to mediate gene repression, this represents an interesting link toward the repressive function of Ogt in ES cells. Another study in mouse ES cells by Shi et al. (2013) further stresses the role of Ogt/Tet1 interaction in gene repression. Besides confirming the interaction of Tet1 and Ogt, they also presented evidence that Tet1 is O-GlcNAcylated. This modification of Tet1 appears to have an effect on its stability. Further results obtained in this study indicate that O-GlcNAcylation positively regulates Tet1 protein levels. On the other hand, loss of O-GlcNAcylation was shown to reduce both Tet1 and 5-hmC levels. Knockdown of Ogt or Tet1 in ES cells appears to result in loss of pluripotency and in an increase in differentiated cells. Data presented in this study implies that Tet1

can recruit repressive complexes, thereby maintaining pluripotency of ES cells. This mechanism might possibly be influenced by O-GlcNAcylation of Tet1. Two further studies indicate an interaction between Ogt and Tet3 (Zhang et al., 2014; Ito et al., 2013). In addition, Zhang et al. (2014) propose that the O-GlcNAcylation of Tet3 regulates its localisation as well as its enzymatic activity. Nonetheless, further studies on the detailed mechanisms by which Tet-mediated repression influences pluripotency are essential at this stage of research.

PcG proteins are known to form multiprotein complexes exhibiting repressive function. Overall, PcG proteins are regulators of gene expression during embryonic development, in stem cells and in somatic cell reprogramming (reviewed by Aloia et al. (2013)). In pluripotent stem cells, PcG complexes have been shown to maintain pluripotency by repressing genes involved in differentiation. Furthermore, the PRC2 component Pcl2 interacts with key members of the pluripotency network (Walker et al., 2010), while Pcl3 was shown to promote self-renewal (Hunkapiller et al., 2012). In addition, PcG complexes play a crucial role in ES cell differentiation. Ezh2, the enzymatically active component of PRC2 is essential for mesodermal differentiation (Shen et al., 2008) and loss of Ring1B, an enzyme in PRC1, hinders the expression of differentiation markers (Leeb and Wutz, 2007). Reprogramming of somatic cells into pluripotent cells has also been shown to be influenced by Polycomb repression. Zhang et al. (2011) reported that overexpression of PRC2 components facilitates reprogramming of into iPS cells. The role of PcG proteins in reprogramming was further underlined by the results of Buganim et al. (2012) who suggest Ezh2 to function as a part of a novel reprogramming transcription factor cocktail.

First results hinting to a role of O-GlcNAc in Polycomb repression where obtained in 2009, when two independent studies showed that the *Drosophila* Ogt enzyme is

encoded by *supersexcombs* (*sxc*) which is a PcG gene (Sinclair et al., 2009; Gambetta et al., 2009). Gambetta et al. (2009) also reported an enrichment of O-GlcNAcylated proteins at Polycomb response elements in *Drosophila*. In 2011, Myers et al. (2011) provided first evidence for the connection between Ogt and PcG proteins in mouse ES cells. Their findings indicate that PRC2 is necessary to maintain normal levels of Ogt and O-GlcNAc distribution. However, this study also indicates no disruption of PRC2 function upon Ogt knockdown. In contrast to the findings of Myers et al. (2011), Tet1-mediated demethylation of CpG islands was shown to be essential for PRC2 recruitment to chromatin (Wu and Zhang, 2011; Aloia et al., 2013). As described above, this process was reported to be influenced by Ogt and O-GlcNAcylation of Tet1. It is therefore surprising, that Ogt knockdown in mouse ES cells does not appear to have any significant effect on PRC2 recruitment and function. A recent study by Maury et al. (2015) further stresses the correlation between O-GlcNAcylation and polycomb-mediated gene repression: Data presented in this study reveals that the catalytic subunit of PRC1, Ring1B, is subject to O-GlcNAcylation at three sites. More importantly, O-GlcNAcylation appears to regulate DNA-binding of Ring1B. In addition, O-GlcNAc modified Ring1B was preferentially detected in proximity of genes related to neural differentiation. Maury et al. (2015) suggest that with a loss of O-GlcNAcylation occuring with progressing differentiation, PRC1 is lost at differentiation-relevant genes providing an on-switch for those genes.

1.2.3. Key reagents for investigating O-GlcNAcylation

Detection methods

Global O-GlcNAcylation can be investigated using conventional immunoblotting techniques. There are two commercially available pan-specific antibodies that are widely used for this purpose: CTD 110.6 (Covance) and RL2 (Sigma) (Comer et al., 2001; Snow, 1987). It is important to note that both antibodies recognize a subset of O-GlcNAcylated proteins. While CTD 110.6 has been demonstrated to detect a wider range of O-GlcNAc modified peptides it also binds to N-GlcNAc(2)-modified glycoproteins (Isono, 2011). Therefore, further investigations using alternative detection methods or upstream sample treatments are generally recommended in order to avoid false positives. Alternative detection methods available for O-GlcNAc are chemoenzymatic or metabolic labelling techniques (Khidekel et al., 2003; Vocadlo et al., 2003).

Enzyme inhibitors

With the aim of studying the function O-GlcNAcylation means of manipulating the O-GlcNAc system are of major importance. This can be achieved by modulation of the two key regulators, Oga and Ogt. Knockdown of both enzymes is possible to a certain degree, but technically challenging and knockdown of Ogt largely impairs cell viability in our hands. Several inhibitors are available for both Ogt and Oga. As inhibition of Ogt also leads to cell death in ES cells (Speakman et al., 2014), modulating the activity of Oga is a useful way of deregulating O-GlcNAcylation and there are several inhibitors commercially available.

PUGNAc is an inhibitor of hexosaminidases that can be used to increase flux through the O-GlcNAc cycle by blocking Oga activity and thereby increasing global O-GlcNAc

levels (Haltiwanger, 1998). In addition to inhibiting Oga, PUGNAc also inactivates lysosomal hexosaminidases due to their structural similarity with Oga (Dorfmueller et al., 2010). Thiamet-G was designed as a selective inhibitor of human Oga that mimics O-GlcNAc (Yuzwa et al., 2008). GlcNAcstatins (GNS) apply a similar approach, but are also characterized by their additional ability to penetrate cells even at nanomolar concentrations thereby minimizing potential off-target effects (Dorfmueller et al., 2010). By targeting a conserved cystein within the catalytically active site of Oga GNS acts as a suicide inhibitor providing an irreversible blockage of the enzyme (Dorfmueller et al., 2010).

1.3. Research objective

This study aims to investigate the function of O-GlcNAcylation in differentiation using different pluripotent stem cell model systems. The role of O-GlcNAc in pluripotent stem cell differentiation has so far been poorly studied, especially regarding human pluripotent stem cells which have become an essential tool for translational and medical research approaches over the last decade. The effects of deregulated O-GlcNAcylation can be studied either by treatment with an Oga inhibitor or by knockdown/overexpression studies of O-GlcNAc-regulating enzymes. Using these tools, O-GlcNAc functions during differentiation into different lineages will be studied. These investigations aim to give hints towards potential effects on lineage commitment in the early mammalian embryo, thereby providing starting points for further studies on embryonic development. In addition, the study will focus on the underlying molecular mechanisms of any possible effects of O-GlcNAc on differentiation. Together, these investigations will contribute significantly to current knowledge on how pluripotent

stem cell differentiation, lineage commitment and early developmental processes are regulated by the posttranslational modification. With a view to the effects of deregulation of O-GlcNAc levels on differentiation, two well-established stem cell models, mouse ES cells and human iPS cells, will be compared. This will provide further understanding of differences and similarities of those model systems.

2. Materials and Methods

2.1. Chemicals

Table (2.1) Chemicals

Chemical	Manufacturer	Order number
2-propanol	Sigma	I9516
4,6-Diamidino-2-phenylindole dihydrochloride	Sigma	D9542
Amersham ECL Western Blotting Detection Reagent	GE Healthcare	RPN2106
Bovine Serum Albumin Fraction V	Sigma	BSAV-RO
Ethanol absolute	VWR	20821.330
Ethylene glycol-bis(2-aminoethylether)-N,N,N',N'-tetraacetic acid	Sigma	E3889
Ethylenediaminetetraacetic acid	Sigma	EDS
Glycine	VWR	101194M
Methanol	VWR	20847.307
milk powder	Marvel	-
Paraformaldehyde	Sigma	158127
Phosphate buffered saline without Ca++ Mg++ or phenol red	Lonza	17-516F
Sodium azide 0.1 M solution	Sigma	8591
Sodium chloride	Sigma	S3014
Sodium dodecyl sulfate	Sigma	L3771
Tris base	Calbiochem	648310-2.5
Triton X-100	Sigma	T8787
Tween 20	Sigma	P9416
UltraPure Agarose	Invitrogen	16500-100

2.2. Cells and general cell culture

2.2.1. Mouse embryonic stem cells

For this study, the mouse ES cell lines E14Tg2aIV and 46C were used. 46C ES cells are an E14Tg2aIV-derived Sox1-GFP knock-in reporter cells (Ying et al., 2003b). All mouse ES cells were cultured on 0.1% gelatin-coated dishes in Glasgow Minimal Essential Medium (GMEM) (Life Technologies) with 10% FCS and LIF (for detailed medium composition see table 2.2). These culture conditions are often referred to as serum/LIF culture and are used to grow mouse ES cells as a mixture of naïve and formative state pluripotent cells. For standard passaging, the cells were dissociated using 0.05% Trypsin-EDTA (Life Technologies) and plated at a density of 40 000 cell/cm^2 on gelatin-coated dishes. Cells were maintained at 37 °C and 5% CO_2.

Table (2.2) GMEM complete

component	supplier	amount/ concentration
Glasgow MEM (BHK-21)	ThermoFisher Scientific	500 ml
Sodium Pyruvate (100 mM)	ThermoFisher Scientific	1 mM
MEM Non-essential Amino Acids Solution (100 ×)	ThermoFisher Scientific	1 ×
2-Mercaptoethanol (≥ 99%)	Sigma	0.1 mM
Leukemia-induced Factor	recombinantly generated	1:1000
Fetal Calf Serum (batch-tested)	obtained from University of Cambridge	10%

2.2.2. Human induced pluripotent stem cells

ChiPS4 human induced pluripotent stem cells used for differentiation experiments were maintained using DEF-CS (Cellartis AB) and provided by Lindsay Davidson,

Human Pluripotent Stem Cell Facility, University of Dundee.

2.2.3. Inhibitors and growth factors

The inhibitors and growth factors given in table 2.3 were added as a direct supplement to cell culture medium.

Table (2.3) Overview inhibitors and chemicals

Name	Function	Supplier	Final concentration
GlcNAcstatin C/G	Oga inhibitor	GlycoBioChem	1 μM
CT99021	GSK-3 inhibitor	Sigma	3 μM (assay-dependent)
PD0325901		Sigma	1 μM
LDN193189	antagonist of BMP receptor isotypes ALK2 and ALK3		100 nM (assay-dependent)
SB431542	inhibitor of ALK4, ALK5 and ALK7		10 μM
Y-27632	ROCK inhibitor		10 μM
Trichostatin A	histone deacetylase inhibitor	Sigma	300 nM
Wnt-C59	PORCN inhibitor		2 μM
Dimethyl sulfoxide	control	Sigma	1:1000
bFGF (recombinant human)	growth factor	Peprotech	30 ng/ml
noggin (recombinant human)	BMP4 antagonist	Peprotech	10 ng/ml
Activin A (recombinant human)		R & D Systems	100 ng/ml
Oncostatin M (recombinant human)	growth and differentiation factor	Preprotech	20 ng/ml
Human HGF	growth factor	Peprotech	10 ng/ml
BMP4 (recombinant human)		Peprotech	assay-dependent

2.2.4. Cell culture coating

Prior to plating of cells, respective culture dishes were coated. For mouse ES cell culture, dishes were coated with 0.1 % gelatin for five minutes at room temperature. For standard culture of undifferentiated human iPS cells dishes were coated with fibronectin for at least one hour at 37°C. Geltrex-coated dishes (at least one hour at 37°C) were used for differentiation assays using human iPS cells. Relevant concentrations and coating reagents are listed in table 2.4.

Table (2.4) Reagents for coating of cell culture dishes

Name	Supplier	Final concentration
Gelatin from porcine skin	Sigma	80 μl of 0.1% gelatin/cm^2
Fibronectin (human plasma)	Millipore	5 μg/cm^2
Geltrex (stem cell-qualified LDEV-free growth factor reduced)	Life Technologies	10 μg/cm^2
Poly-l-ornithine	Sigma	15 μg/cm^2
Laminin	Life Technologies	1 μg/cm^2

2.2.5. Differentation assays

Spontaneous differentiation of mouse ES cells in embryoid bodies

Upon LIF withdrawal mouse ES cells differentiate spontaneously. In this assay, cells were cultured in multicellular aggregates termed embryoid bodies (EBs) and medium without LIF which allows them to differentiate into multiple lineages.

Mouse ES cells (46C or E14Tg2aIV) were detached enzymatically using StemPro Accutase (Thermo Fisher Scientific) described and resuspended in GMEM (Life Technologies) with 10% FCS (GMEM/FCS) but without LIF and with either GNS G (1 μM) or DMSO (1:1000). The cell suspension was adjusted to a concentration of 1×10^5

cell/ml. Cells were then plated as hanging drops of 25 μl volume onto the lid of a petri dish. In order to avoid evaporation the petri dish was filled with 1× PBS and the lid with the hanging drops was placed on top. The hanging drop culture was incubated for three days to allow the cells to form EBs. Subsequently, EBs were either harvested for gene expression analysis in RLT lysis buffer or transferred to a non-adherent petri dish with GMEM/FCS and GNS G/DMSO. Cells were then cultured for another 2 days before they were lysed in RLT lysis buffer for RT-qPCR (reverse transcriptase quantitative polymerase chain reaction).

Neural differentiation of ChiPS4 (high-density protocol)

For neural differentiation, we applied a dual SMAD inhibition protocol as described by Chambers et al. (2009). Before differentiation was initiated, ChiPS4 cells were plated on geltrex-coated dishes (10 μg/cm^2) at a density of 4 $\times$ 10^4 cells per cm^2. The cells were grown until they formed a uniform monolayer (approximately 3 d). During this time medium was changed daily. Differentiation was then initiated by changing the culture medium to serum replacement medium (Knockout DMEM, 20 % knockout serum replacement, 1x glutamax, 1x non-essential amino acids, 100 M 2-mercaptoethanol, all from Life Technologies) supplemented with 10 μM SB431542 and 100 nM LDN193189. Cells were kept under these conditions with daily medium changes for 5 days. On day 5 of the differentiation medium was changed to 75% Knockout serum replacement (KOSR)/25% neurobasal medium with 1 $\times$ N2, 1 $\times$ B27 and LDN193189. The medium was subsequently changed to increasing amounts of neurobasal/N2B27 (50%, 75% and 100%) every other day while maintaining 100 nM of the inhibitor LDN193189. Protein and ribonucleic acid (RNA) lysates were obtained at various timepoints throughout the differentiation.

Neural differentation of ChiPS4 in embryoid bodies

Neural differentiation of ChiPS4 in EBs was performed based on the protocol described by Koch et al. (2009). In order to generate EBs, a cell suspension with a concentration of 2×10^5 cells/ml in DEF base medium with bFGF (30 ng/ml), noggin (10 ng/ml) and Y-27632 (10 μM) was prepared. This equates to 10,000 cells per 50 μl drop. The cell suspension was then dispensed onto a non-adhesive 96-well v-bottom plate (50 μl per well). Plates were centrifuged at 500 $\times$ g for 5 min. This leads to the formation of cell aggregates in the bottom of the wells which will then form EBs within the next 48 h.

EBs were then transferred into 10 cm dishes with knockout Dulbecco's minimal essential medium (DMEM) containing SB-431542 (10 μM) and LDN-193189 (0.5 μM) (Day 1 of differentiation). Over the course of the coming days the medium was sequentially changed from knockout DMEM to neurobasal/N2B27: Every other day, half of the medium was replaced by neurobasal/N2B27 while the inhibitors were maintained at the previously stated concentrations. On day 9 of differentiation the medium was changed completely to neurobasal/N2B27/SB-431542/LDN-193189 and the EBs were plated out onto poly-l-ornithine (15 μg/cm^2 in sterile water) and laminin (1 μg/cm^2 in phosphate buffered saline (PBS)) coated dishes. The medium was then changed every 48 h for another 6 days.

On day 15 of the differentiation neuroectodermal islands (three-dimensional cell aggregates with neural-tube-like structures) can be isolated in order to remove potential contamination with other lineages. For this, neural islands were dislodged using a p1000 pipette. The cells were then transferred to a 50 ml tube and centrifuged at 300 $\times$ g for 2 min. Neural islands were resuspended in fresh neurobasal/N2B27/SB-

431542/LDN-193189 medium and plated on matrigel (20 μg/cm^2) coated dishes. Cells were maintained for another 9 days with media changes every other day. On day 24 of the differentiation cells were fixed and analysed by immunofluorescence.

Hepatocyte differentiation of ChiPS4

Human iPS cell were differentiated into a hepatocyte-like cell type using a protocol which was based on Medine et al. (2011). This three-step protocol starts with definitive endoderm differentiation during the first four days. This is followed by a hepatic differentiation step (five additional days) and a hepatic maturation (nine days).

Cells were initially plated at a density of 4×10^4 cells per cm^2 on matrigel (10 μg/cm^2) coated plastics. When cells had grown to confluence, differentiation was initiated by changing the medium to RPMI/B27 (RPMI-1640 with 1 $\times$ B27, 1 $\times$ Glutamax and penicillin/streptomycin, all from Thermo Fisher Scientific) supplemented with Activin A (100 ng/ml) and Wnt3a (50 ng/ml). Medium was changed every 25 hours for the next three days. On day four of differentiation, the medium was changed to serum replacement medium (Knockout DMEM, 20 % knockout serum replacement, 1x glutamax, 1x non-essential amino acids, 100 mM 2-mercaptoethanol, all from Life Technologies) supplemented with 1% dimethylsulfoxide (DMSO). Cell were kept in this medium for an additional five days with medium changes every 48 hours. On day eight of the differentiation, medium was changed to L-15 maturation medium (Leibovitz L-15 medium with tryptose phosphate broth (final concentration 8.3%, Sigma), FCS (final concentration 8.3 %), 10 μM hydrocortisone 21-hemisuccinate, Insulin (1 μM), 1 $\times$ Glutamax, 0.2% ascorbic acid, from ThermoFisher Scientific unless otherwise stated) supplemented with hepatocyte growth factor (HGF) (10 ng/ml) and oncostatin M (20 ng/ml). Medium was again renewed every other day for another

nine days. Protein and RNA lysates were obtained at various timepoints throughout the differentiation.

Cardiomyocyte differentiation of ChiPS4

Cardiomyocyte differentiation of ChiPS4 was carried out according to the protocol described by Burridge et al. (2014). First, cells plated in DEF medium (including noggin (10 ng/ml), bFGF (30 ng/ml) and Y-27632 (10 μM)) at a density of 50,000 cells/cm^2 on matrigel (10 μg/cm^2) coated plastics. Medium was changed the next day to DEF medium without Y-27632. Approximately 48 h after plating medium was changed to RPMI/B27-ins with CT99021 (5 μM) (D0 of differentiation). Cells were kept in this medium for two days. On day 2 of the differentiation medium was changed to RPMI/B27-ins with Wnt-C59 (2 μM). The medium was changed to RPMI/B27-ins on day 4 of the assay and cells were maintained in this medium with medium changes every other day. The cells were kept until day 10 of cardiomyocyte differentiation. Lysates for RNA extraction and Western Blot were generated at relevant timepoints during this assay.

2.2.6. BMP pathway stimulation

ChiPS4 were plated at a density of 30 000 cells/cm^2 on geltrex-coated dishes. The cells were kept in standard culture conditions (DEF medium supplemented with bFGF and noggin) including relevant drugs (i.e. DMSO, GNS) for three days. Medium was then changed to DMEM/FCS (DMEM, 10% FCS, 1 $\times$ glutamax) and different amounts of BMP4 were added (between 0.1 to 100 ng/ml). After overnight stimulation with BMP4, RNA was harvested.

2.2.7. RNAi

siRNA-induced knockdown in mouse ES cells

For small inhibitory RNA (siRNA)-induced knockdown of Oga or Ogt cells were plated into 6-well dishes at a concentration of 2×10^5 cells/well. While the cells were still in suspension, they were reversely transfected with siRNAs targeting either Oga or Ogt (50 pmol per well) (ON-TARGETplus SMARTpool against mouse Ogt or mouse Mgea5; Thermo Fisher Scientific). As a control, cells were transfected with a non-targeting siRNA pool (50 pmol per well) (ON-TARGETplus non-targeting pool; Thermo Fisher Scientific). For each well both siRNAs and 5 μl of the transfection reagent Lipofectamine RNAiMAX (Life Technologies) were diluted in 250 μl Opti-MEM (Thermo Fisher Scientific) each. The two dilutions were then combined and incubated for 5 min before they were added to the cells. Using the same protocol cells were transfected again after 24 h. Lysates for RNA purification and Western blotting were obtained after another 24 h in culture.

2.2.8. Transfections

For overexpression of full length Oga and a shorter splicing variant, E14 mouse ES cells were transfected with the plasmids pCAG-IP-Oga-full length or pCAG-IP-Oga-splicing variant respectively. One million cells were plated in 6 cm dishes and reversely transfected using 4.5 μg DNA and 9 μl Lipofectamine 2000 (Life Technologies) in 1000 μl OptiMEM (Life Technologies). After 24 hours, the transfection was repeated. Cells were then incubated for another day before being harvested respective to the desired subsequent analysis method.

2.3. Molecular biology methods

2.3.1. DNA extraction

For some of the methods described in the following sections genomic DNA was extracted. As a first step, medium was removed and cells were washed with PBS before being harvested in 250 μl DNA lysis buffer (75 mM NaCl, 25 mM Ethylenediaminetetraacetic acid (EDTA), 1% sodium dodecyl sulfate (SDS)). RNAse A (Sigma) was added at a final concentration of 50 μg/ml and samples were incubated for 30 minutes at 37 °C Subsequently, 10 μl of proteinase A (10 mg/ml stock, Sigma) was added and incubation was continued at 55 °C for one hour. To each sample 365 μl chloroform and 78.5 μl 5 M sodium chloride were added. The samples were then incubated on a rotating wheel for 45 minutes. Samples were centrifuged at 16 000 $\times$ g for 10 minutes. The upper aqueous phase was transferred to a new tube with 250 μl isopropanol. Samples were mixed, incubated for five minutes and then centrifuged again for another twelve minutes. After this, a pellet had formed at the bottom of the tube. The supernatant was discarded and pellets were dried for approximately 15 minutes. The DNA pellet was then dissolved in an appropriate volume of Tris-EDTA buffer.

2.3.2. Real time quantitative PCR

Reverse transcription

In order to generate cDNA (complementary DNA), RNA was reversely transcribed using the qScript cDNA synthesis kit (Quanta Biosciences) according to the manufacturer's instructions. Reverse transcription was performed on 1000-500 ng of RNA. The temperature profile for the reverse transcription reaction is depicted in table 2.5.

Table (2.5) Temperature profile for cDNA synthesis

22 °C	5 min
42 °C	30 min
85 °C	5 min

2.3.3. RNA extraction

Cells were harvested in 350 μl RLT buffer supplemented with 3.5 μl beta-mercaptoethanol (Qiagen). Subsequently, RNA extraction including on-column DNase digest was performed using the NucleoSpin RNA kit (Macherey-Nagel) according to the manufacturer's instructions. RNA was eluted in RNase-free water (40-50 μl) and RNA concentrations were determined using a Nanodrop 1000 Spectrophotometer (Thermo Scientific).

Real-time PCR

Firstly, cDNA was diluted to a total volume of 80 μl with nuclease-free water. Reaction mixes were prepared by combining SooFast EvaGreen Supermix (BioRad), nuclease-free water and primers according to table 2.6. Mastermix was then pipetted into the respective wells and 3 μl of diluted cDNA were added per well. The plate was sealed and spun down. Subsequently, the reaction was performed using a CFX384 Touch Real-Time PCR Detection System (BioRad). The temperature profile used for this reaction is depicted in table 2.7. All reactions were performed as technical triplicates. Relevant primer sequences for human cells are listed in table 2.8.

Table (2.8) RT-qPCR primers, human

primer name	sequence (5' to 3')	primer bank ID/ publication
beta-actin, forward	CATGTACGTTGCTATCCAGGC	4501885a1

beta-actin, reverse	CTCCTTAATGTCACGCACGAT	4501885a1
Oct4, forward	CTTGAATCCCGAATGGAAAGGG	4505967a1
Oct4, reverse	GTGTATATCCCAGGGTGATCCTC	4505967a1
Nanog, forward	AAGGTCCCGGTCAAGAAACAG	153945815c2
Nanog, reverse	CTTCTGCGTCACACCATTGC	153945815c2
Pax6, forward	AGTGCCCGTCCATCTTTGC	189083679c2
Pax6, reverse	CGCTTGGTATGTTATCGTTGGT	189083679c2
Sox1, forward	ATGCACCGCTACGACATGG	D'Amour et al. (2005)
Sox1, reverse	CTCATGTAGCCCTGCGAGTTG	D'Amour et al. (2005)
Pax7, forward	ACCCCTGCCTAACCACATC	207029245c1
Pax7, reverse	GCGGCAAAGAATCTTGGAGAC	207029245c1
FoxG1, forward	GAGCGACGACGTGTTCATC	375151583c1
FoxG1, reverse	GCCGTTGTAACTCAAAGTGCTG	375151583c1
Sox17, forward	CTCCGGTGTGAATCTCCCC	145275218c2
Sox17, reverse	CACGTCAGGATAGTTGCAGTAAT	145275218c2
FoxA2, forward	GGAGCAGCTACTATGCAGAGC	194363755c1
FoxA2, reverse	CGTGTTCATGCCGTTCATCC	194363755c1
AFP, forward	CTTTGGGCTGCTCGCTATGA	4501988c1
AFP, reverse	GCATGTTGATTTAACAAGCTGCT	4501988c1
albumin, forward	TGCAACTCTTCGTGAAACCTATG	215982788c1
albumin, reverse	ACATCAACCTCTGGTCTCACC	215982788c1
Brachyury, forward	TATGAGCCTCGAATCCACATAGT	19743811c1
Brachyury, reverse	CCTCGTTCTGATAAGCAGTCAC	19743811c1
GATA5, forward	CTTCGTGTCCGACTTCTTGGA	157952217c1
GATA5, reverse	CCGAGGCATTCCTTGTGGA	157952217c1
cTNT, forward	ATGATGCATTTTGGGGGTTA	Fonoudi et al. (2013)
cTNT, reverse	CAGCACCTTCCTCCTCTCAG	Fonoudi et al. (2013)
Hand1, forward	GAGAGCATTAACAGCGCATTCG	196259809c2
Hand1, reverse	CGCAGAGTCTTGATCTTGGAGAG	196259809c2
Klf2, forward	CTACACCAAGAGTTCGCATCTG	49574523c2
Klf2, reverse	CCGTGTGCTTTCGGTAGTG	49574523c2
FoxJ1, forward	GTGCTTCATCAAAGTGCCTCG	260764002c2
FoxJ1, reverse	GCCTCGGTATTCACCGTCA	260764002c2
MAPK12, forward	AGTGGCTTTTACCGCCAGG	48255969c1
MAPK12, reverse	GACTGGAAGGGCCGATACAG	48255969c1
Klf8, forward	GTCAGTCTGCCAAATAAGATGGG	226530794c2
Klf8, reverse	ATGGAGGTGGGGTCAACTTTC	226530794c2
Smad7, forward	TTCCTCCGCTGAAACAGGG	299890804c1
Smad7, reverse	CCTCCCAGTATGCCACCAC	299890804c1
TGFBI, forward	CACTCTCAAACCTTTACGAGACC	170650698c1

TGFBI, reverse	CGTTGCTAGGGGCGAAGATG	170650698c1
SEMA3A, forward	GTGCCAAGGCTGAAATTATCCT	100913215c1
SEMA3A, reverse	CCCACTTGCATTCATCTCTTCT	100913215c1
BMPR1B, forward	ATTTGCAGCACAGACGGATATT	377823722c2
BMPR1B, reverse	GAGGCAGTGTAGGGTGTAGGT	377823722c2
GDNF, forward	GCAGACCCATCGCCTTTGAT	299473777c2
GDNF, reverse	CCACACCTTTTAGCGGAATGC	299473777c2
Cdx2, forward	GACGTGAGCATGTACCCTAGC	372624391c1
Cdx2, reverse	GCGTAGCCATTCCAGTCCT	372624391c1

Quantification

Relative quantification was performed according to the method described by Pfaffl (2001). Beta-actin was used as reference gene.

2.3.4. Cloning of Oga overexpression plasmids

With the aim of generating plasmids for overexpression of Oga, cDNA was generated from RNA extracted from E14 mouse ES cells according to the procedure described in section 2.3.3 and section 2.3.2. The relevant sequences for cloning were then amplified by polymerase chain reaction (PCR) using the primers in table 2.10. For amplification, PrimeSTAR HS DNA Polymerase (Clonetech) was used according to the manufacturer's instructions.

The amplified sequences were then cloned into the pCAG-IP vector (Niwa et al., 2002) using the XhoI and NotI restriction sites. Plasmid preparations were performed using the QIAprep Spin Miniprep Kit (Qiagen) or the PureLink HiPure Plasmid Filter Midiprep Kit (ThermoFisher Scientific).

Maps of the generated plasmids are depicted in figure 2.1 and figure 2.2.

Table (2.6) Reaction for RT-qPCR

5 μl	SooFast EvaGreen Supermix
1 μl	primer mix (10 pmol/μl each)
1 μl	nuclease-free water
3 μl	cDNA
10 μl	total volume

Table (2.7) Temperature profile for RT-qPCR

95 °C	5 min	
95 °C	15 sec	
57 °C	32 sec	40 ×
65-95 °C	5 s	melt curve, increment 0.5 °C

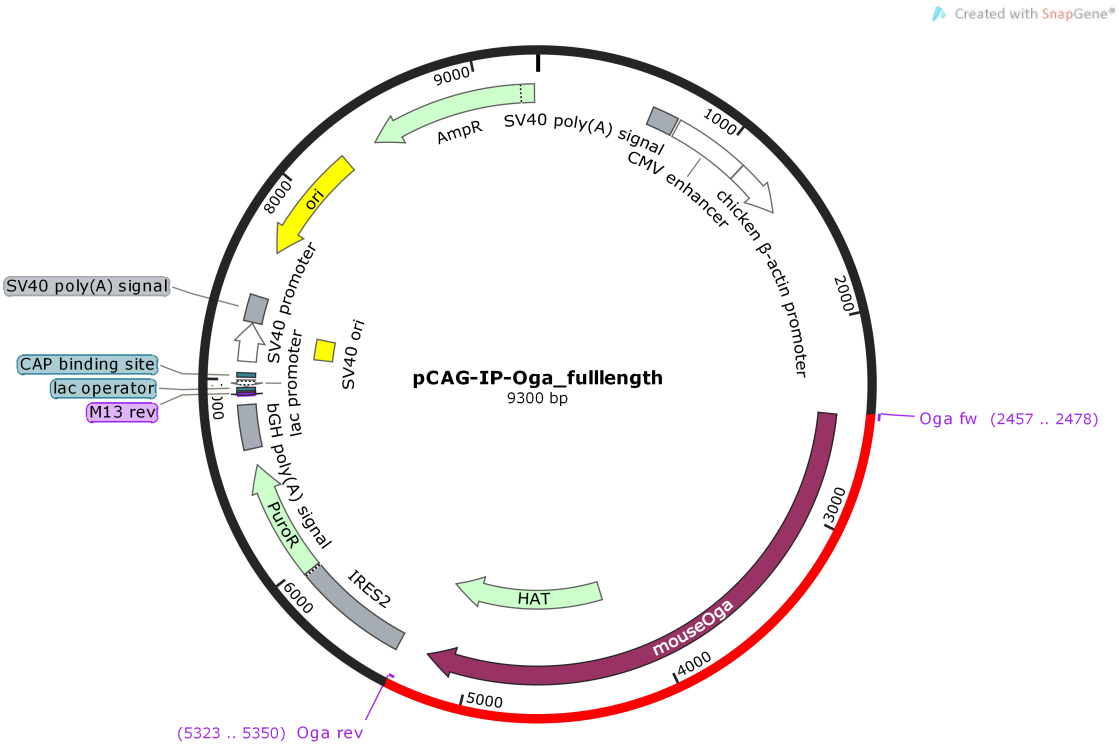

Figure (2.1) pCAG-IP-Oga full length

Table (2.9) RT-qPCR primers, mouse

primer name	sequence (5' to 3')	primer bank ID/ publication
beta-actin, forward	GGCTGTATTCCCCTCCATCG	6671509a1
beta-actin, reverse	CCAGTTGGTAACAATGCCATGT	6671509a1
Ogt, forward	GACGCAACCAAACTTTGCAGT	20982829a1
Ogt, reverse	TCAAGGGTGACAGCCTTTTCA	20982829a1
MGEA5, forward	GGGTTATGGAGCAGAGAAAAGAG	15011884a1
MGEA5, reverse	CCTGGCGAAATAGCATAGATGAA	15011884a1
BMP2, forward	GGGACCCGCTGTCTTCTAGT	6680794a1
BMP2, reverse	TCAACTCAAATTCGCTGAGGAC	6680794a1
Eomesodermin, forward	GGCCCCTATGGCTCAAATTCC	258645095c1
Eomesodermin, reverse	CCTGCCCTGTTTGGTGATG	258645095c1
Brachyury, forward	GCTTCAAGGAGCTAACTAACGAG	6678203a1
Brachyury, reverse	CCAGCAAGAAAGAGTACATGGC	6678203a1
FGF5, forward	AAGTAGCGCGACGTTTTCTTC	3721900a1
FGF5, reverse	CTGGAAACTGCTATGTTCCGAG	3721900a1
Tdpoz3, forward	TCTCCAATGTCCAATGCTTTCTG	157057187c1
Tdpoz3, reverse	ACGGTTCCAACTATGCTCACC	157057187c1
Zfp352, forward	AAGTCCCACATCTGAAGAAACAC	23821036a1
Zfp352, reverse	GGGTATGAGGATTCACCCACA	23821036a1
Zscan4, forward	CAGATGCCAGTAGACACCAC	-
Zscan4, reverse	GTAGATGTTCCTTGACTTGC	-
AthoI, forward	GAGTGGGCTGAGGTAAAAGAGT	6680742a1
AthoI, reverse	GGTCGGTGCTATCCAGGAG	6680742a1
Sp110, forward	ATGAAGGTGAACATCGCCTATG	30425106a1
Sp110, reverse	GGACAGAGGGACCAGATTTTG	30425106a1

Table (2.10) Oga cloning primers

primer name	sequence (5' to 3')
Oga full length, forward	CTCGAGCGGCAGGAGGATGGTG
Oga full length, reverse	GCGGCCGCAGTCAACTGGCTGATCTGCT
Oga, splicing variant, forward	CTCGAGCGGCAGGAGGATGGTG
Oga, splicing variant, reverse	CTAGTGCGGCCGCCTACAGCAAACGCT↲ GGAACTCTC

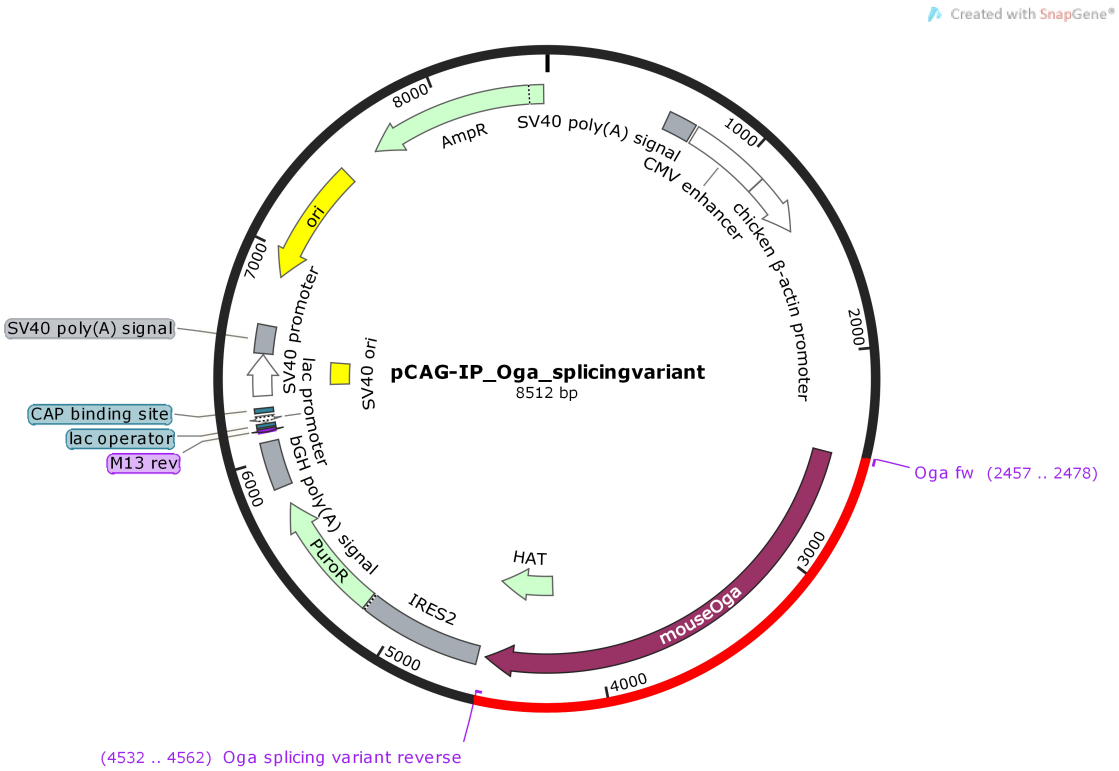

Figure (2.2) pCAG-IP-Oga splicing variant

2.4. Protein biochemistry

2.4.1. Western Blot

Lysate generation

In order to generate whole cell lysates, cells were first washed twice with 1 × PBS. An appropriate volume of lysis buffer (20 mM Tris, pH 7.5; 150 mM NaCl; 1 mM EDTA; 1 mM EGTA; 1 μM GNS G; cOmplete Tablets, Mini, EDTA-free (Roche)) was then added to the cells and they were scraped off using a cell scraper. The lysate was transferred to a tube and centrifuged in a tabletop centrifuge at maximum speed for ten minutes. Subsequently, the supernatant was transferred to a new tube.

Bradford assay

Total protein concentration was determined using a photometric assay according to Bradford (1976). For each sample, 2 μl of lysate were pipetted into a 96-well plate and 200 μl of Bradford reagent Pierce Coomassie (Bradford) Protein Assay Kit (ThermoFisher Scientific) were added to the sample. Absorption at 625 nm was determined after brief incubation using a Multiskan Spectrum (ThermoFisher Scientific) plate reader.

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)

Three parts of protein lysate were mixed with one part 4 × NuPAGE LDS Sample Buffer (ThermoFisher Scientific). Samples were then boiled for five minutes. Both for SDS-PAGE and transfer Novex NuPAGE (ThermoFisher Scientific) tanks were used. Novex NuPAGE Tris-Glycine gels were washed and inserted into the tank which was

then filled with 1 × NuPAGE MOPS SDS Running Buffer (ThermoFisher Scientific). SeeBlue Plus2 Prestained Standard (Thermo Fisher Scientific) (5 μl for a gels with 1 mm thickness) and samples were loaded (5-15 μg total protein). Gels were run at 100-200 V at constant voltage.

Transfer

Subsequently, proteins were transferred to a Polyvinylidene difluoride (PVDF) membrane (Immobilon-P, Millipore). For this, PVDF membranes were first activated by incubation in methanol. Filter paper and sponges were soaked in transfer buffer (table 2.11). In order to prepare the transfer sandwich, a soaked sponge was placed into the transfer cassette first, followed by filter paper, the gel, the membrane, a second filter paper and finally a second sponge. The cassette was then inserted into the tank and filled with transfer buffer. The rest of the tank was filled with ice and cold water. As a last step the transfer was performed at 50-45 V for 1.5 h.

Table (2.11) Transfer buffer

2.9 g	glycine
5.8 g	Tris base
0.37 g	SDS
ad 800 ml	water
200 ml	methanol

Antibody incubation and development

After the transfer, the membrane was incubated in blocking buffer (TBS-T (table 2.12) with either 5 % milk powder, 3% BSA or 1 % BSA corresponding to the blocking buffer the antibody was diluted in). Subsequently, the membrane was incubated in primary

antibody dilution overnight at 4 °C The next day, the membrane was washed three times with TBS-T for approximately five minutes. Secondary antibodies were diluted in 5 % milk in TBS-T according to table 2.14 and incubated for 45 minutes at room temperature. The membrane was then washed again as previously described. To develop membranes that had been incubated with a HRP-coupled secondary antibodies ECL was used according to the manufacturer's instruction. Blots were then developed using Amersham Hyperfilm and an autodeveloper. When fluorochrome-coupled antibodies had been used, detection was performed using a Odyssey CLX scanner (LI-COR Biotechnology) and Image Studio (LI-COR Biotechnology) software.

Table (2.12) TBS-T

10 mM	Tris-HCl (pH 7.3)
100 mM	NaCl
0.1 %	Tween 20

Table (2.13) Primary antibodies for Western Blot

target	supplier	product number	dilution	blocking/dilution in
beta-actin	Abcam	ab6276	1:10000	3 % BSA
O-GlcNAc (CTD110.6)	Covance	MMS-248R	1:5000	1 % BSA
Oga	Abcam	ab124807	1:1000	5 % milk
Ogt	Abcam	ab50271	1:500	3 % BSA

2.4.2. Immunofluorescence staining and analysis

In order to prepare cells for immunofluorescence staining, medium was aspirated and cell were washed twice with 1 × PBS. Cells were then fixed using a 4 % formaldehyde solution (paraformaldehyde solved in 1 × PBS) for 20 minutes at room temperature. Subsequently, cell were again washed with 1 × PBS.

Table (2.14) Secondary antibodies for Western Blot

Antibody	supplier	product number	dilution
Goat anti-mouse IgM DyLight 680	ThermoFisher Scientific	SA5-10154	1:15000
Anti-mouse IgG (H+L), DyLight 680	Cell Signaling Technology	5470	1:15000
Donkey anti-mouse IgG, HRP conjugated	Millipore	AP192P	1:10000
Goat anti-mouse IgM - mu chain, HRP conjugated	Abcam	ab97230	1:5000
Anti-rabbit IgG, HRP conjugated	Cell Signalling Technology	7074	1:1000

Cells that would be stained with antibodies against Oct4, Nanog or HNF-III-β were treated with 90 % ice-cold methanol for five minutes on ice before permeabilisation/blocking. For combined permeabilisation and blocking, cells were then treated with 4% BSA in PBS/0.2% Triton-X-100. Cells that would subsequently be stained with antibodies for detecting extracellular markers (TRA-1-60, TRA-1-80) were not permeabilized and blocked using 4% BSA in PBS without any detergent. Cells were then washed again with 1 $\times$ PBS and primary antibodies were incubated overnight at 4 °C All primary antibody dilutions were made up in 4 % BSA/PBS according to table 2.15.

After primary antibody incubation cells were washed four times with 1 $\times$ PBS (approximately ten minutes per wash). Secondary antibody dilutions were then prepared in 4 % BSA/PBS (see table 2.16, all antibodies diluted 1:500) and incubated for 20 minutes on the cells at room temperature (protected from light). For nuclear staining DAPI (4',6-diamidino-2-phenylindole) (final concentration: 300 nM) was added to the secondary antibody dilution.

Subsequently, secondary antibodies were taken off and the cells were again washed

Table (2.15) Primary antibodies for immunofluorescence

target	supplier	product number	dilution
Oct-3/4 Antibody (C-10)	Santa Cruz	sc-5279	1:100
Nanog (D73G4) XP® Rabbit mAb	Cell Signaling Technology	4903	1:200
TRA-1-60 antibody	Santa Cruz	sc-21705	1:50
TRA-1-81 (TRA-1-80) antibody	Santa Cruz	sc-21706	1:50
HNF-3-β antibody (MC-20)	Santa Cruz	sc-6554	1:100
ZO-1 / TJP1 Antibody	ThermoFisher Scientific	40-2200	1:300
N-cadherin / CDH2 Antibody (NCD-2)	ThermoFisher Scientific	13-2100	1:300
Monoclonal Anti-β-Tubulin III (neuronal) antibody produced in mouse	Sigma	T8578	1:600
Sox2 (D6D9) XP Rabbit mAb	Cell Signaling Technology	3579	1:400

Table (2.16) Secondary antibodies for immunofluorescence

Antibody	supplier	product number
Goat anti-mouse IgG / IgM (H+L), Alexa Fluor 488	Thermo Fisher Scientific	A-10680
Donkey anti-rabbit, Alexa Fluor 488	Thermo Fisher Scientific	A-12206
Donkey anti-mouse, Alexa Fluor 568	Thermo Fisher Scientific	A-10037
Donkey anti-mouse, Alexa Fluor 647	Thermo Fisher Scientific	A-31571
Donkey anti-Rat IgG (H+L), Alexa Fluor 488	Thermo Fisher Scientific	A-21208
Donkey anti-goat IgG (H+L), Alexa Fluor 488	Thermo Fisher Scientific	A-11055

four times with 1 × PBS. Cells were then analyzed using a DeltaVision Microscopy Imaging System (GE Healthcare) and softWoRkx (AppliedPrecision).

2.5. Epigenetic methods

2.5.1. Histone extraction

Histone extracts were generated for as starting material for analysis of histone modifications, i. e. by Western Blot or enzyme-linked immunosorbent assay (ELISA) (see below). First, cells were detached with accutase (Life Technologies) and resuspended in 1 × PBS. The cell suspension was spun down at 1000 rpm and 4°C for five minutes. The supernatant was discarded, the cell pellet was resuspended in TEB buffer (10^7 cells per 1 ml buffer; PBS + 0.5 % Triton + 0.02 % sodium azide + Pefabloc SC (0.1 mM, Roche)) and samples were left to lyse on ice for ten minutes with gentle stirring. Subsequently, samples were centrifuged (3000 rpm for five minutes for volumes over 1.5 ml, 10 000 rpm for 1 minute for volumes up to 1.5 ml). The supernatant was removed and the pellet was resuspended in extraction buffer (200 μl buffer per 10^7 cells; 0.5 N hydrochloric acid + 10 % gylcerol) and incubated on ice for 30 minutes. Samples were then centrifuged again at 12 000 rpm at 4°C for five minutes. The supernatant was transferred to a new vial and acetone was added (600 μl per 10^7 cells). After mixing the samples they were incubated overnight at -20°C.

The next day, samples were centrifuged at 12 000 rpm for five minutes, the supernatant was discarded and the pellets were air-dried. As a last step the pellet was dissolved in distilled water (30-50 μl per 10^7 cells).

2.5.2. Histone 3 acetylation: enzyme-linked immunosorbent assay

Analysis of histone 3 acetylation levels by ELISA was performed using the Histone H3 Acetylation Assay Kit (Abcam: ab115102) according to the manufacturer's instructions. Colorimetrc readout was performed using a Multiskan Spectrum (ThermoFisher Scientific) plate reader.

2.5.3. Bisulfite sequencing

DNA methlyation was analysed by bisulfite sequencing. Genomic DNA was extracted from E14 mouse ES cells according to the procedure for DNA extraction described above (section 2.3.1). Subsequently, DNA samples were bisulfite converted using the Epitect Bisulfite Conversion kit (Qiagen) according to the manufacturer's instructions. For each sample, 2 μg of total DNA were converted.

Bisulfite-converted DNA (1 μl of the conversion mix) was then amplified using Epi-Taq HS polymerase (Clonetech) according to the manufacturer's standard protocol and primers designed for the amplification of bisulfite-converted templates (table 2.17, designed using MethPrimer by Li Lab, Peking Union Medical College Hospital). Following amplification, 25 μl of the PCR reaction were treated with 6 U Klenow fragment for 20 minutes at room temperature. Samples were then cleaned up using the Qiaquick Gel Extraction kit (Qiagen) according to the kit instructions from step four onwards. DNA was eluted in 25 μl EB buffer.

PCR products were then cloned into a pSC-B vector. For this, the Stratclone Blunt Cloning (Agilent Technologies) kit was used according to the respective instructions. Transformed cells were plated onto X-Gal- and ampicillin-containing LB agar

Table (2.17) bisulfite-specific primers

primer name	sequence (5' to 3')
beta-actin, forward	GTTTTTTTTGTTTTTTGAGTTTGG
beta-actin, reverse	TCACCCACATAAAAATCCTTCTAAC
REST, forward	AAAAGAAAGAAGGATGTTGAGGTTTA
REST, reverse	AAAATCTAAAAACACCCCCACTTA
Sp110, forward	TGAATGTAGTTAAGGAGAGGTATAGTT
Sp110, reverse	ATCAAAAAAAACCACAAATAAATAC
Klf4, forward	TGTTTTTTTATGTGTAAGGTAAGGTG
Klf4, revers	CTTCTAACAATAACTTCAACAACCC
Xist, forward	GTAATTGGTTGTTTTTATTTAGTT
Xist, reverse	CCTAAAATATCTATACCTCTTATTTAC
Rex1, forward	TATTTTGGTATTTTGGATTGGTTTT
Rex1, reverse	AATCACTTTCTCACCATAAACTCTC

plates. Per reaction, five white colonies were picked for plasmid preparation. Plasmid minipreps were performed using the QIAprep Spin Miniprep Kit (Qiagen). Plasmids were then send for sequencing using M13 forward and reverse primers. Analysis of the sequencing results was performed using BiQ Analyzer v2.0 (Bock et al., 2005).

2.5.4. Sample preparation for HPLC-MS

High-performance liquid chromatography-mass spectrometry (HPLC-MS) analysis of genomic DNA for E14 and ChiPS4 cells was performed by Dr Rachel Amouroux in the group of Petra Hajkova (Imperial College, London). Genomic DNA was extracted using the ZR-Duet DNA/RNA miniprep kit form Zymo Research (Cambridge) according to the manufacturer's instructions. DNA was eluted in LC-MS grade water (Fisher Scientific, order number W/0112/17).

2.6. Flow Cytometry

Flow cytometry was performed using an LSRFortessa cytometer (BD Biosciences, Belgium). Data analysis was performed using FlowJo (FlowJo, USA).

Viability assay

To determine viability of live cells by flow cytometry approximately 2×10^6 cells were detached using 1× TrypLE Select Enzyme (ThermoFisher Scientific, USA). The enzyme was inhibited by adding standard culture medium (at least four times the volume of the TrypLE Select volume). The cell suspension was subsequently transferred to 5 ml polystyrene round-bottom tubes (Falcon, USA) and cells were pelleted by centrifugation (200× g, 2 min). After discarding the supernatant the cells were resuspended in 300 μl 5 μg/ml DAPI (Sigma-Aldrich, USA) in PBS containing 1 % FCS (PBS/1% FCS). Subsequently, the cells were analysed for DAPI staining as an indicator of dead cells.

Cell cycle analysis

For analysing cell cycle, approximately 2×10^6 cells were detached by the method described above. Again, the cells were transferred into standard culture medium in 5 ml polystyrene round-bottom tubes (Falcon, USA) and pelleted by centrifugation (200× g, 2 min). The supernatant was discarded and cells were resuspended in the small volume remaining. Per tube, 1 ml of PBS/1% FBS was added. Cells were pelleted again by centrifugation. Subsequently, the supernatant was poured off and the pellet was resuspended in the volume remaining in the tube. While vortexing, 1 ml of ice-cold 90% methanol was added to the cells. The cells were then incubated

in the fixative for at leat 30 min at room temperature. After this, the cells were washed twice with 1 ml PBS/1% FBS. Following the last centrifugation step, the supernatant was removed completely and cells were resuspended in 300 μl of staining buffer (50 μg/ml propidium iodide (PI) (Sigma-Aldrich, USA), 50 μg/ml RNase A (Sigma-Aldrich, USA) in PBS/1% FBS). The cell suspension was then incubated for 20 min at room temperature protected from light. Analysis of propidium iodide staining was performed by flow cytometry and cell cycle analysis was carried out using FlowJo with a Watson Pragmatic algorithm.

Analysis for RFP or EGFP expression

In order to analyse the expression of RFP or EGFP of relevant reporter cell lines, cells were dissociated and stained with DAPI using the method described above. The cells were then analysed for their fluorescence properties and DAPI was used to exclude dead cells from the analysis.

2.7. RNA sequencing

2.7.1. Cell treatments and sample preparation

ChiPS4 were plated in 6-well-plates at a density of 30 000 cells/cm^2 and treated with 1 mM GNS or the equivalent volume of DMSO. Cells were kept under standard culture conditions for three days and RNA lysates were then generated. RNA extraction was performed as described in section 2.3.3 and the RNA was handed over to the Genetics Core Services Unit at Ninewells Hospital and Medical School in Dundee where ribosomal depletion, library preparation and sequencing were carried out.

2.7.2. Bioinformatic analysis

Bioinformatic analysis was performed in collaboration with Pieta Schofield (Data Analysis Group, University of Dundee).

Alignment against the human genome reference (GRCh38) was carried out using STAR2.5 (Dobin et al., 2013). Differential gene expression analysis was then performed using the Bioconductor package edgeR (Robinson et al., 2010). The Bioconductor packages biomaRt (Durinck et al., 2009) and StringDB (Franceschini et al., 2013) were used for further analysis of the data, i.e. gene ontology and kegg pathway analysis. The relevant R code sections are provided in the appendix (section A.1.3).

3. O-GlcNAc function in pluripotent stem cells and differentiation

3.1. Effects of GlcNAcstatin on mouse embryonic stem cells

3.1.1. GlcNAcstatin G treatment does not affect viability or cell cycle

GlcNAcstatins have been reported to inhibit Oga in a highly specific manner by binding to the active site of the enzyme and preventing substrate glycosylation. Furthermore, these inhibitors have been shown to induce hyper-O-GlcNAcylation in human cell lines. (Dorfmueller et al., 2006, 2009) In order to investigate the effects of increased global levels of O-GlcNAc in pluripotent stem cells, GlcNAcstatin G (GNS G) was added to the culture medium at a concentration of 1 μM.

With the aim of ruling out potential off-target effects of GNS G on pluripotent stem cells, I set out to validate that cell viability and cell cycle are unaffected by treatment with GNS G. This was essential as effects on viability or cell cycle might impact on the cells ability to differentiate into different lineages. Therefore, flow cytometry experiments on E14Tg2aIV (E14) mouse ES cells were carried out.

In order to determine potential effects on viability cells were treated with GNS G for approximately 48 h and stained with DAPI. This allows determination of the number of dead cells within the population by inclusion of the dye. The findings of these flow cytometry experiments are illustrated in Figure 3.1 A and B. The overall percentage of dead cells was low (less than 8 %) and no significant differences could be

detected between GNS G treated cells and those that had been treated with a DMSO vehicle control. It is evident from these results that GNS G treatment does not impair viability of E14 mouse ES cells.

Previously obtained data on mouse 46C ES cell proliferation suggest that the cell cycle of mouse ES cells is also not affected by GNS G treatment (Speakman et al., 2014). To extend and validate those findings, cell cycle analysis using PI staining and flow cytometry was performed. This method allows a precise determination of the number of cells in G1-, S- and G2-phase by assessing the cell's DNA content. Again, cells were treated for at least 48 h with GNS G before being submitted to the analysis. It becomes clear from Figure 3.1 C, D and E that GNS G has no effect on the cell cycle. Taken together these results are consistent with our previous findings showing that GNS G does not impair cell proliferation.

It is apparent from the results described above that GNS G does not exhibit any unintended effect on the viability and cell cycle of mouse ES when used at a final concentration of 1 μM. It can therefore be concluded that the observed effects described below are unrelated to any impact of GNS G treatment on proliferation or viability.

3.1.2. Inhibition of Oga by GlcNAcstatin increases global O-GlcNAc levels and leads to a feedback upregulation of Oga

To further characterize the effects of GNS treatment on mouse ES cells, previously reported effects of GNS C on global O-GlcNAcylation as well as on protein levels of Oga and Ogt (Speakman et al., 2014) had to be validated. With the aim of describing the effects of GNS treatment on O-GlcNAc levels and the two regulators, Western Blotting was performed using protein lysates obtained from E14 mouse embryonic

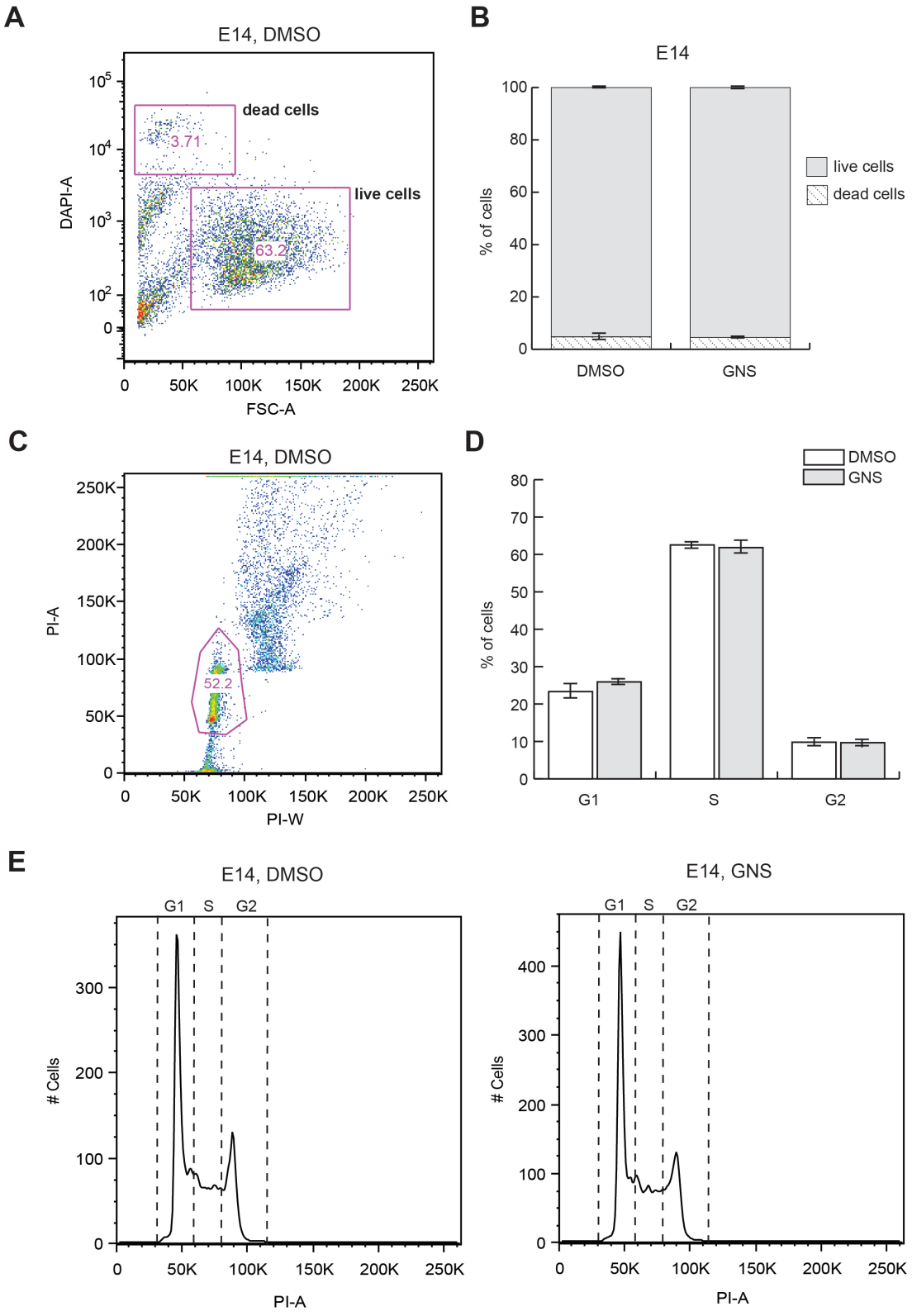

Figure (3.1) GNS G treatment does not affect viability or cell cycle of E14Tg2aIV mouse ES cells. **A, B:** Flow cytometry of DAPI stained cells showing that viability is not affected by GNS G treatment. **A**: Gating strategy excluding debris and cell duplets. Numbers in the boxes indicate percentage of total counts. **B**: Percentages of single cells after exclusion of debris and duplets with high DAPI (dead cells) and low DAPI (live cells) staining. **C, D:** Flow cytometry of propidium iodide stained cells indicate no changes in cell cycle. **E:** Representative cell cycle profiles of control treated (left graph) and GNS G treated (right) cells. (n=3 independent biological repeats for all experiments, errorbars represent standard error of the mean (SEM))

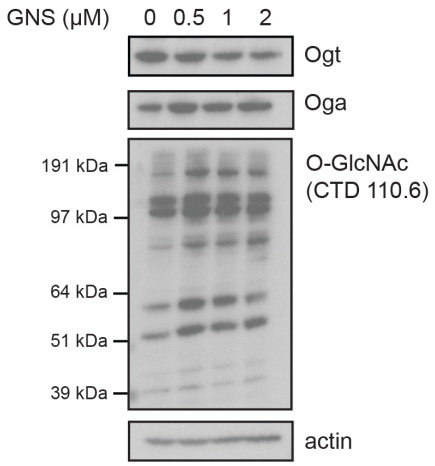

Figure (3.2) GNS G increases O-GlcNAcylation and leads to feedback regulation of Oga and Ogt in mouse ES cells. A: Whole cell lysates obtained from E14 mouse ES cells treated with various concentrations of GNS G show increased O-GlcNAcylation, feedback upregulation of Oga and downregulation of Ogt protein levels. A representative Western Blot is shown.

stem cells treated with GNS G at three different concentrations (0.5 μM, 1 μM and 2 μM) for 48 h. As shown in Figure 3.2, treatment with GNS G dramatically increases global O-GlcNAc levels even at 0.5 μM. In addition, inhibition of Oga by GNS G could be demonstrated to affect total protein levels of Oga and Ogt. While Ogt protein levels decrease with increasing GNS G concentration, Oga protein levels become elevated. These findings demonstrate a feedback regulation of the two enzymes by global O-GlcNAcylation levels. It can also be seen from this result that a further increase in GNS concentration from 1 μM to 2 μM only has little effect on global O-GlcNAcylation as well as on Ogt and Oga proteins levels.

Those results are in good agreement with the effects of Oga inhibitors described for other models as well as with previous data from our group (Speakman et al., 2014). As expected, treatment with GNS G leads to elevated O-GlcNAc levels and feedback regulation of Oga and Ogt. However, it is important to note that modulation of this

system is only possible to a certain extent and the addition of GNS at concentrations above 1 μM does not affect O-GlcNAcylation any further.

3.1.3. Elevated O-GlcNAc levels delay spontaneous differentiation into mesodermal and endodermal lineage

The key feature of pluripotent stem cells is their ability to differentiate into cells of all three germ layers as well as the germline. In order to understand the function of O-GlcNAcylation in early development and pluripotent stem cells it is essential to elucidate potential roles of the modification in mouse ES cell differentiation. Overall very little work has been done on this subject. Kim et al. (2009) describe a progressive reduction of O-GlcNAc levels and Ogt protein during cardiomyocyte differentiation. Moreover, their results indicate that excessive O-GlcNAcylation in response to treatment with the Oga inhibitor PUGNAc impairs cardiogenesis. In accordance to this study, recent results confirm that O-GlcNAc levels decrease during mouse ES cell differentiation (Speakman et al., 2014). In addition, a delay in the onset of neural differentiation has been shown in response to Oga inhibition by GNS C (Speakman et al., 2014). In order to extend previous findings the effects of increased protein O-GlcNAcylation on the ability of mouse ES cells to differentiate spontaneously in embryoid bodies were investigated. In contrast to lineage-specific differentiation protocols, this assay allows the cells to differentiate into cells of any lineage rather than directing differentiation towards a specific fate by changes of culture medium or addition of growth factors. Due to this undirected nature of spontaneous differentiation the resulting cell population contains cells of all three lineages which allows the assessment of potential effects on each lineage by observing the expression of specific marker genes (Martin et al., 1977; Coucouvanis and Martin, 1995). For this assay 46C

cells were cultured in embryoid bodies using standard culture medium without LIF, but supplemented with either GNS C or DMSO. Cells were differentiated for five days and quantitative RT-PCR for marker genes was performed on mRNA extracted from undifferentiated cells as well as from cells at day 3 and day 5 of differentiation. Figure 3.3 A, B and C illustrate the results of the analysis for the endoderm marker *BMP2* and the mesoderm markers *Eomesodermin* and *Brachyury*. The expression of all three genes is reduced in GNS treated cells when compared to cells treated with a vehicle control. This indicates a delay in differentiation into both endodermal and mesodermal lineage. Combined with previously obtained data on ectoderm differentiation (Speakman et al., 2014) these results clearly indicate a crucial role of O-GlcNAcylation during mouse ES cell differentiation which appears to be lineage-independent.

Surprisingly, the expression of the transient epiblast marker *Fgf5* appears to be unchanged by GNS treatment (Figure 3.3 D). This finding implies that while lineage specification is impaired by GNS treatment transition from naïve to formative pluripotency remains unaffected. Further experiments have also demonstrated that expression of naïve pluripotency markers like *Nanog*, *Rex1 (Zfp42)*, *Esrrb* and *Dppa3* are unaltered by GNS induced Oga inhibition supporting the conclusion that naïve to formative transition is not modulated by O-GlcNAc levels (Speakman et al., 2014).

These data clearly demonstrate that elevated O-GlcNAc levels affect differentiation of mouse ES cells in a lineage-independant manner. The data also implies that early onset of differentiation remains largely unaffected as both the marker expression levels at early timepoints in differentiation (day three) and expression of *Fgf5* remain unchanged by GNS treatment.

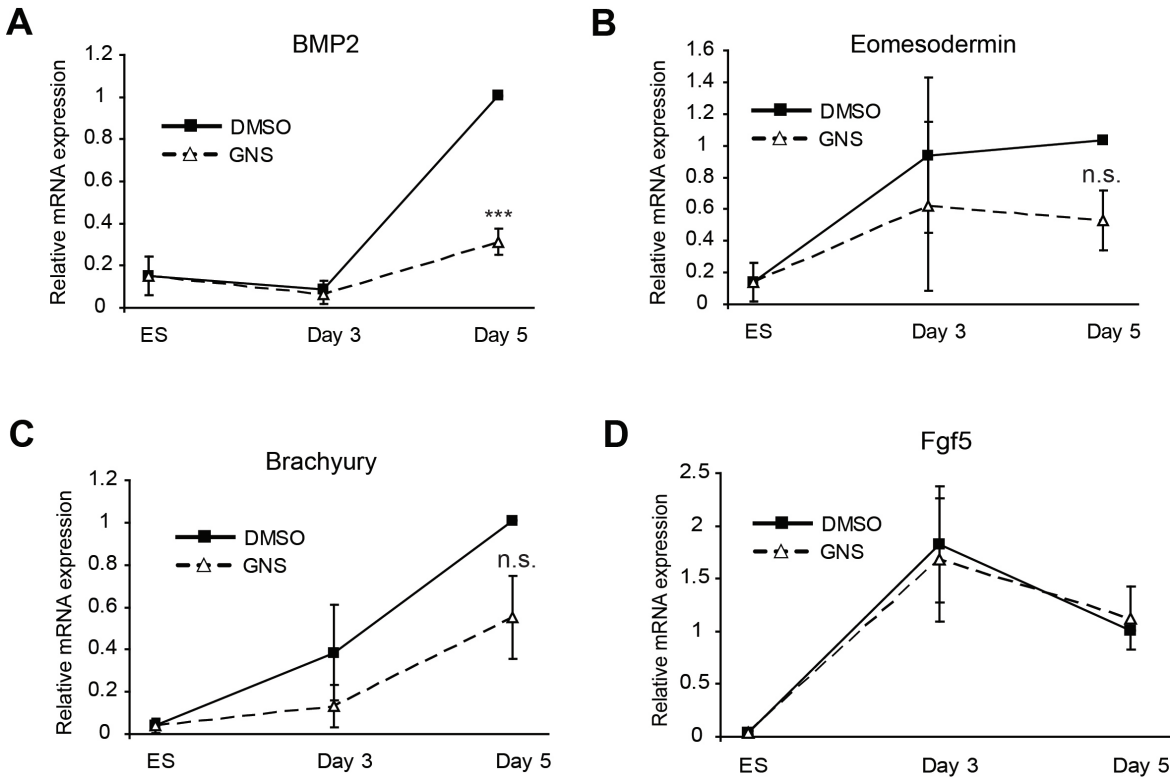

Figure (3.3) Endoderm and mesoderm differentiation is hindered by GNS C treatment while naïve to formative transition is unchanged. RT-qPCR analysis of marker genes for endoderm (**A**), mesoderm (**B, C**) and the epiblast (**D**) during spontaneous differentiation in embryoid bodies. n=3 independent biological repeats, errorbars indicate SEM; ***: $p \leq 0.001$; **: $p \leq 0.01$; *: $p \leq 0.05$; n.s.: $p > 0.05$ (Student's t-test, unpaired)

3.2. Effects of GlcNAcstatin on human iPS cells

3.2.1. GlcNAcstatin G treatment does not affect viability or cell cycle

Human iPS cells represent an excellent system to study differentiation in a human pluripotent cell type without the need of using ES cells and a multitude of targeted differentiation assays for those cells have been developed over the last years (Takahashi et al., 2007; Taura et al., 2009; Hirami et al., 2009; Tanaka et al., 2009). In order to extend the findings in mouse ES cells to another model system, ChiPS4 human iPS cells generated by the human pluripotent stem cell facility of the University of Dundee were used for the experiments described in the following.

In order to ensure that GNS treatment would not interfere with cell viability or cell cycle a similar set of flow cytometry experiments as described in section 3.1.1 were performed. In contrast to the experiments done with mouse ES cells, human iPS cells were treated with GNS G for three days instead of 48 h. The reason for this prolonged treatment is the fact that human iPS cells divide slower than mouse ES cells and potential effects could otherwise be masked by a too brief treatment.

Again, cells were analysed by flow cytometry for inclusion of DAPI in order to determine the number of viable cells. The gating strategy and the result of this experiment are depicted in Figure 3.4 A and B. As can be seen from Figure 3.4 B, the percentage of living cells was very high (more than 80 % of single cells) and there were no detectable differences between cells that had been treated with GNS and the vehicle control. It can therefore be concluded that GNS G does not affect human iPS cell viability.

With the aim of ruling out any possible effects of GNS on the cell cycle of ChiPS4

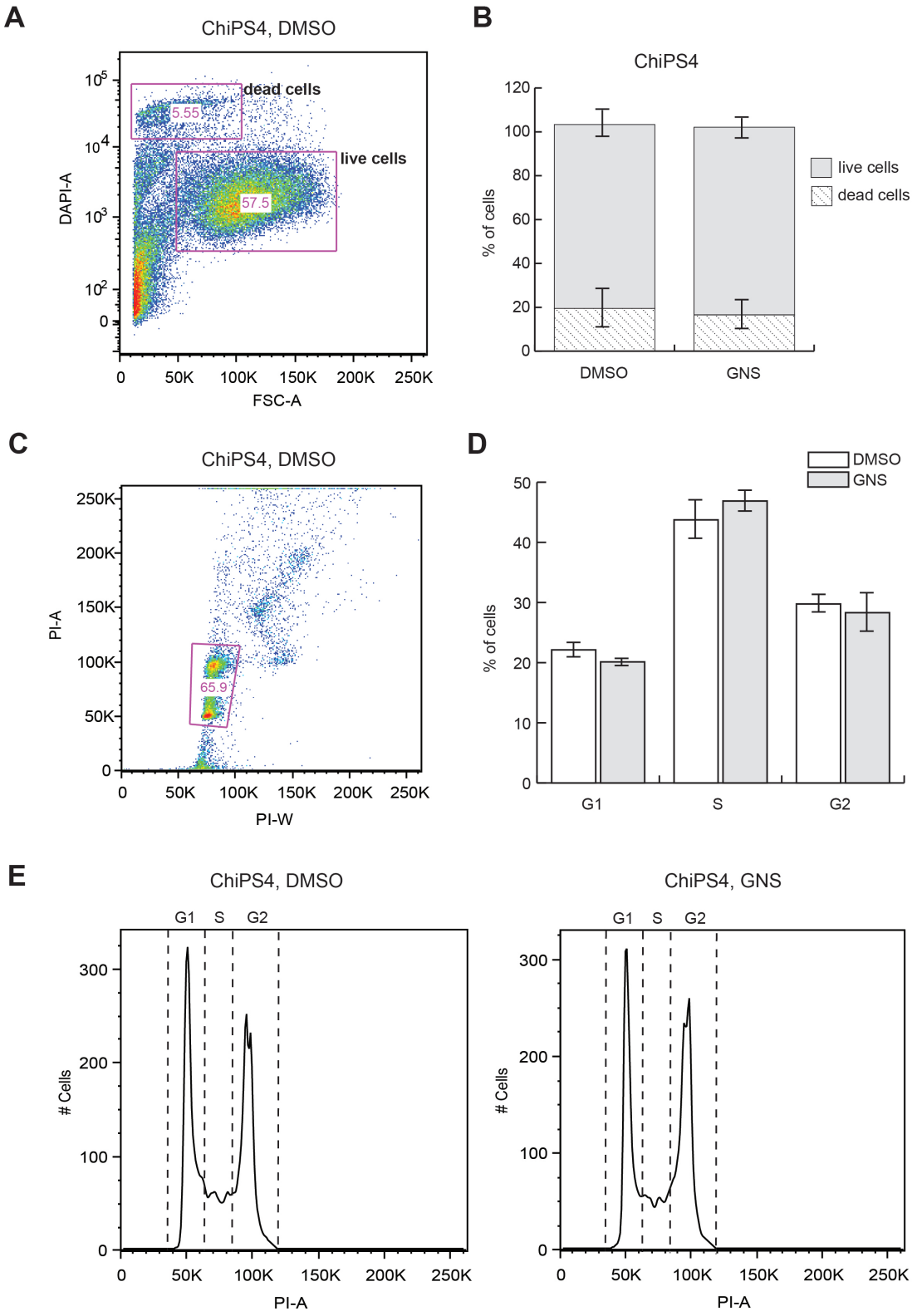

Figure (3.4) GNS G treatment does not affect viability or cell cycle of ChiPS4 cells. A, B: Flow cytometry of DAPI stained cells showing that viability is not affected by GNS G treatment. **A**: Gating strategy excluding debris and cell duplets. Numbers in the boxes indicate percentage of total counts. **B**: Percentages of single cells after exclusion of debris and duplets with high DAPI (dead cells) and low DAPI (live cells) staining. **C, D:** Flow cytometry of propidium iodide stained cells indicate no changes in cell cycle. **E:** Representative cell cycle profiles of control treated (left graph) and GNS G treated (right) cells. (n=3 independent biological repeats for all experiments, errorbars represent SEM)

cells we performed PI staining for DNA content on living cells followed by flow cytometry analysis (gating strategy in Figure 3.4). The results of this experiment are illustrated in Figure 3.4 D and E. As can be seen from the cell cycle profiles there were no differences between ChiPS4 treated with GNS and those cells that had been treated with a DMSO control. Interestingly, when comparing the cell cycle profiles of ChiPS4 with our previous experiment using mouse ES cells (see Figure 3.1 D, E) it is clear that there are more ChiPS4 cells in G2 than mouse ES cells. This increase in the G2-population for human iPS cells corresponds to their slower cell cycle. To summarize, GNS G does not have any detectable effect on either human iPS cell viability or cell cycle. Any effects of GNS treatment on differentiation are therefore unlikely to be caused by cell death or reduced proliferation of the cells.

3.2.2. Inhibition of Oga by GlcNAcstatin increases global O-GlcNAc levels and leads to a feedback regulation of Oga and Ogt

As described in section 3.1.2 GNS treatment leads to an increase in global O-GlcNAc levels and respective feedback regulation of Oga and Ogt in mouse ES cells. As to ensure that these effects can be reproduced in human iPS cells, ChiPS4 were treated with two different concentrations of GNS G for three days and whole cell lysates were harvested and analysed by Western Blot. As illustrated in Figure 3.5 O-GlcNAcylation was found to be significantly increased at 1 μM GNS. In addition, Ogt protein levels were dramatically decreased in GNS treated cells due to the expected feedback mechanism. Interestingly, the amount of Oga protein did not seem to be as strongly affected. It is also important to note that an increase in GNS concentration to 2 μM did not elevate O-GlcNAc levels any further and had very little additional effect on

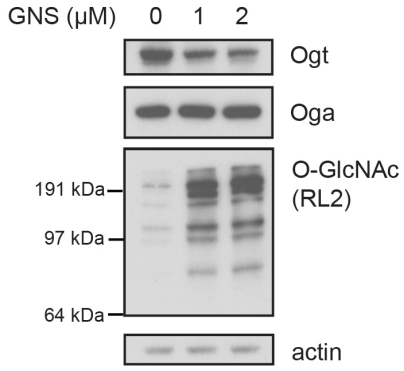

Figure (3.5) GNS G increases O-GlcNAcylation and leads to feedback regulation of Oga and Ogt in human iPS cells. A: Whole cell lysates obtained from ChiPS4 cells treated with various concentrations of GNS G show increased O-GlcNAcylation, feedback upregulation of Oga and downregulation of Ogt protein levels. A representative Western Blot is shown.

Oga and Ogt protein levels. These findings are consistent with our results in mouse ES cells and demonstrate clearly that GNS G concentrations of beyond 1 μM do not have an additional effect in modulating O-GlcNAcylation.

3.2.3. GNS does not affect pluripotent stem cell markers

Data by Maury et al. (2013) had suggested that while excess O-GlcNAcylation can modify differentiation of human ES cells it does not affect these cells in their undifferentiated state. In order to determine whether ChiPS4 can be maintained in an undifferentiated state when treated with GNS G, cells were analysed for their expression of pluripotency markers. As illustrated in Figure 3.6 A gene expression of Nanog and Oct4 remain unchanged in ChiPS4 that had been treated with GNS G for three days. In order to further validate those findings, immunofluorescence staining and analysis was performed on ChiPS4 (again after three days of treatment with GNS). It can be observed that both the transcription factor Nanog (Figure 3.6 B) and the extracellular stem cell marker TRA-1-60 (Figure 3.6 C) are expressed in the

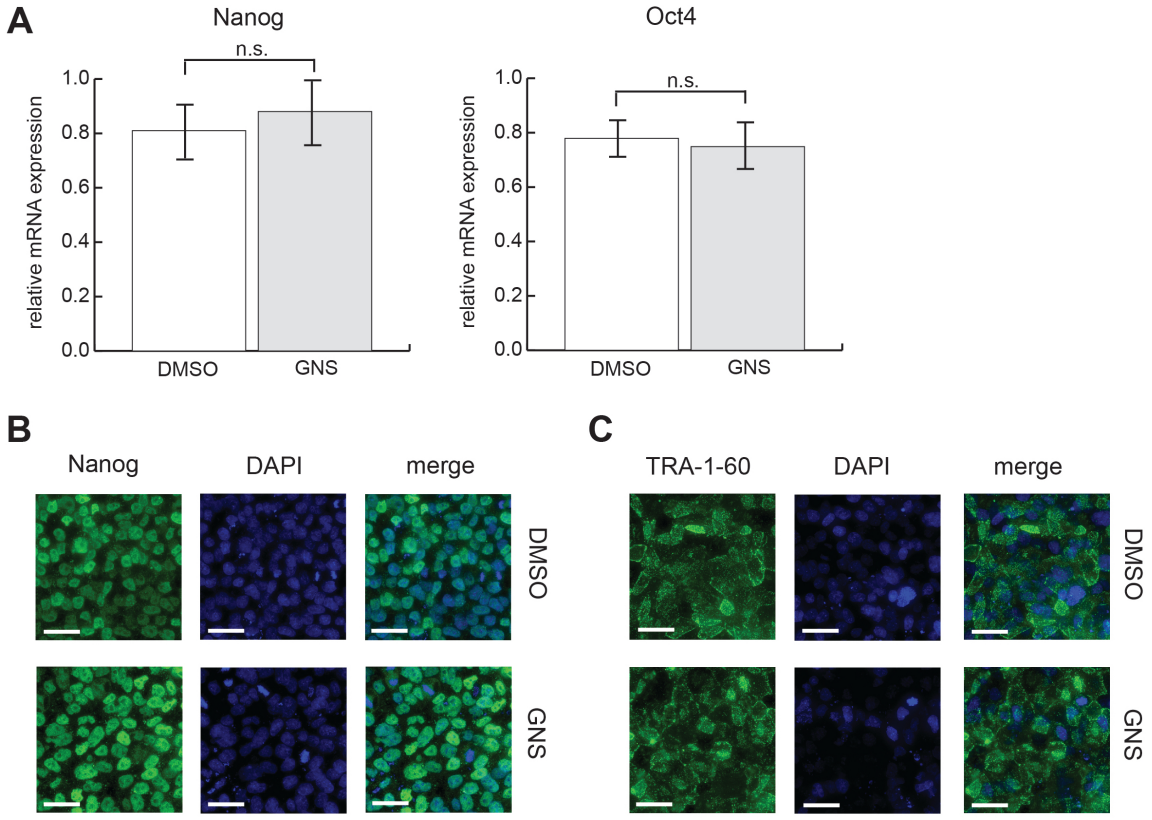

Figure (3.6) GNS G treatment does not affect pluripotent stem cell markers in ChiPS4. **A**: Gene expression of *Nanog* and *Oct4* remain unaffected by GNS G treatment (n=6 independent biological replicates, errorbars represent SEM; n.s.: $p > 0.05$ (Student's t-test, paired).). **B, C**: Pattern and intensity of staining for the transcription factor Nanog and the extracellular stem cell marker TRA-1-60 are unaffected by GNS G treatment. Representative images are shown. scalebar: 40 μm

same pattern and quantity when comparing GNS treated ChiPS4 to the respective control. These data indicate that pluripotency markers of ChiPS4 is not affected by high O-GlcNAcylation and that undifferentiated cells are maintained under GNS G treatment.

3.3. The role of O-GlcNAc during differentiation of human iPS cells

Data presented in the previous sections provides validation of the effects of GNS treatment on global O-GlcNAc levels and their regulating enzymes in undifferentiated pluripotent stem cells. I then set out to investigate potential effects of elevated O-GlcNAcylation on differentiation of human iPS cells into different lineages. Therefore, one directed differentiation assay per lineage was chosen and the dynamics of global O-GlcNAc levels as well as expression of lineage marker genes were investigated with the aim of determining potential consequences of GNS treatment.

3.3.1. O-GlcNAc functions in hepatocyte differentiation of ChiPS4

In order to investigate the differentiation into endoderm an assay for generation of hepatocytes from human iPS cells was chosen based on the protocol described by Medine et al. (2011). A brief overview of the three main steps of this protocol is illustrated in Figure 3.7 A. Cells are first differentiated into definitive endoderm which is achieved by supplementing the medium with Activin A and the GSK-3 inhibitor CT99021. After three days all cells are expected to have committed to endodermal lineage and can be further differentiated into hepatocytes in two following steps: hepatocyte differentiation and hepatic maturation (Medine et al., 2011).

Initially, lysates for RNA extraction were generated at different timepoints during the hepatocyte differentiation which were then analysed for the expression of relevant marker genes by RT-qPCR. Lysates were obtained from both cells treated with GNS and control cells treated with DMSO. A set of markers consisting of pluripotency markers (*Oct4, Nanog*), definitive endoderm markers (*Sox17, FoxA2*) and hepatocyte

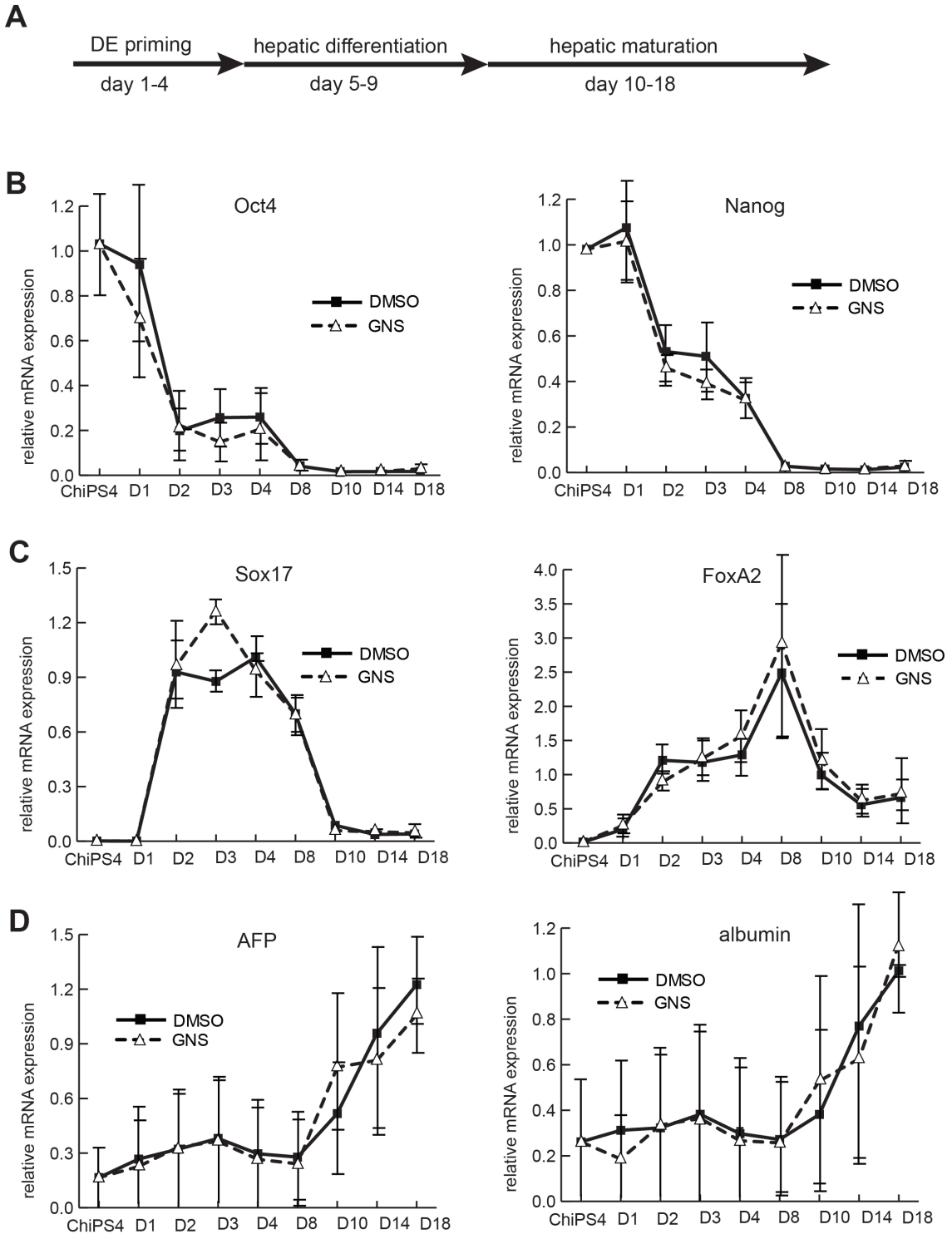

Figure (3.7) RT-qPCR analysis of marker gene expression during hepatocyte differentiation A: Time-course of the hepatocyte differentiation protocol; **B**: mRNA levels of the pluripotency markers *Oct4* and *Nanog*; **C**: gene expression profiles of the definitive endoderm markers *Sox17* and *FoxA2*; **D**: expression of the hepatocyte marker genes *alpha-fetoprotein (AFP)* and *albumin*. n=3 independent biological repeats, errorbars represent SEM

markers (*alpha-fetoprotein (AFP), albumin*) was chosen in order to obtain detailed information about the differentiation process. The results of the RT-qPCR analysis are shown in Figure 3.7 B-D. As expected, the expression of the pluripotency markers *Oct4* and *Nanog* drops very quickly with the onset of differentiation within the first two days of this assay (Figure 3.7 B). For both pluripotency markers there are no differences in gene expression when comparing GNS treated and untreated cells. Figure 3.7 C illustrates the timecourse of the expression of the transient markers *Sox17* and *FoxA2*. Both genes are transiently expressed during the definitive endoderm priming step (day 2 to day 8) and experience a sharp drop in expression as cells progress towards a hepatic fate. Again, there are no significant differences between cells that had been treated with GNS and the vehicle control indicating that differentiation onset and lineage specification to endoderm are not affected by Oga inhibition. For the two mature hepatocyte markers *AFP* and *albumin* mRNA levels begin to increase steadily from day 10 onwards as cells begin to complete differentiation into hepatocytes (see Figure 3.7 D). Interestingly, high variations in the magnitude of expression of these genes between experiments were observed. These are most likely caused by biological differences between individual experiments which will impact on the cells' ability to fully differentiate into mature hepatocytes (i.e. plating density). It can be summarized from the data presented above that both lineage commitment to endoderm and hepatic maturation remain unaffected by Oga inhibition in response to GNS treatment in this assay.

With the aim of further validating these observations, the expression of marker proteins for definitive endoderm differentiation was investigated. Therefore, the same differentiation assay as previously described was applied but the experiment was stopped after completion of the definitive endoderm priming step (day 4). Cells were then fixed

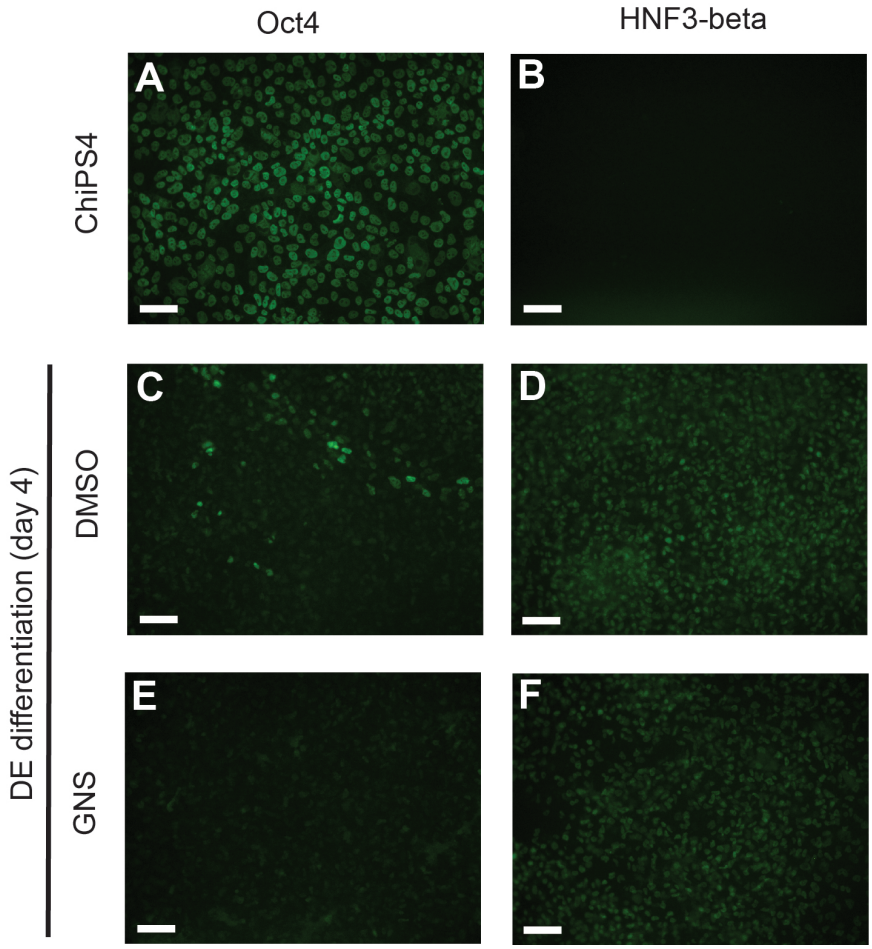

Figure (3.8) Immunofluorescence staining for markers of pluripotency (Oct4) and definitive endoderm differentiation (HNF3-β). Oct4 (**A**) and HNF3-β (**B**) expression in undifferentiated ChiPS4; **C**, **D**: Expression of Oct4 and HNF3-β after 4 days of differentiation in DMSO treated cells; **E**, **F**: Expression of Oct4 and HNF3-β after 4 days of differentiation in GNS treated cells. Representative images are shown. Scalebar: 50 μm.

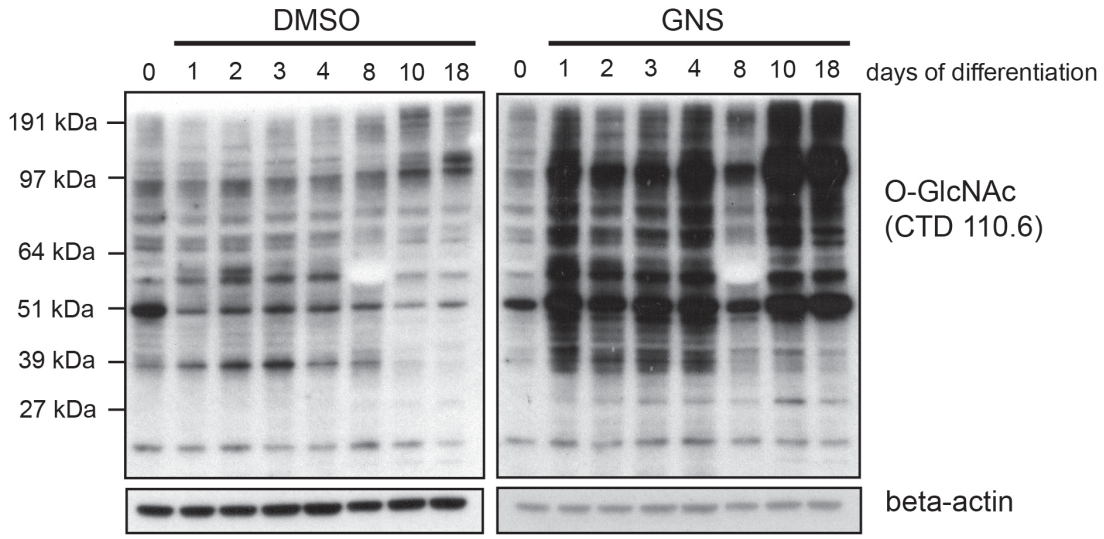

Figure (3.9) Variations in global O-GlcNAc levels during hepatocyte differentiation of ChiPS4. Left panel: O-GlcNAcylation of ChiPS4 treated with DMSO. Right panel: O-GlcNAcylation of GNS treated cells. A representative Western Blot is shown.

and subjected to immunofluorescence staining and microscopy. The samples were then analysed for the expression of Oct4 as a marker of pluripotency and HNF3-β as endoderm marker. The respective results are presented in Figure 3.8. As expected Oct4 was highly expressed in undifferentiated ChiPS4 cells, whereas HNF3-β was not expressed (Figure 3.8 A and B). After four days of differentiation into definitive endoderm only few cells remained Oct4-positive in the DMSO control (Figure 3.8 C). For cells that had been treated with GNS Oct4 expression was reduced even more than in the vehicle control (Figure 3.8 E). However, both DMSO and GNS treated cells displayed high levels and a uniform expression pattern of HNF3-β after definitive endoderm differentiation (Figure 3.8 D and F). These findings underline the previous results and demonstrate that GNS treated cells successfully differentiate into definitive endoderm.

In addition to the analysis of marker expression it is important to understand how O-GlcNAc levels change over the course of differentiation. For this experiment whole

cell lysate was harvested at different timepoints of the differentiation and Western Blot analysis was performed using an antibody that detects O-GlcNAc. As illustrated in Figure 3.9 the overall levels of O-GlcNAcylation change little over the course of this assay. Strikingly, O-GlcNAcylation exhibits a very dynamic pattern which is most likely caused by the modification being removed from certain proteins and attached to others. At day 10 and 18 of the assay a strong shift of O-GlcNAcylation from proteins of low molecular weight towards those with high molecular weight can be observed. At those timepoints only very few proteins below a molecular weight of 51 kDa appear to be weakly modified. The right panel of Figure 3.9 displays the respective analysis of O-GlcNAcylation for GNS treated cells. In comparison to the vehicle control O-GlcNAc levels are clearly increased. Interestingly, the levels appear to be fluctuating more in the GNS treated cells than in the DMSO control. It appears that O-GlcNAcylation is reduced for day two and three of the differentiation followed by an increase on day four. On day eight O-GlcNAcylation drops again dramatically whereas both day 10 and 18 are characterized by very high O-GlcNAc levels. However, the pattern of O-GlcNAcylation remains dynamic with a shift towards more modification of high molecular weight proteins at the final stage of the differentiation assay. Taken together, the O-GlcNAcylation pattern and levels are dynamic during hepatocyte differentiation. These findings imply that while the levels of the modification appear to be of minor importance for successful differentiation, the addition and removal of O-GlcNAc from certain proteins might be essential as the pattern of O-GlcNAcylation is maintained even in GNS treated cells.

3.3.2. O-GlcNAc functions in cardiomyocyte differentiation of ChiPS4

With the aim of investigating potential effects of elevated O-GlcNAc levels on ChiPS4 differentiation into mesodermal lineage a cardiomyocyte differentiation protocol was applied. In brief, ChiPS4 were differentiated according to the protocol described by Burridge et al. (2014) and the Wnt signalling pathway was modulated in order to trigger the lineage commitment. Again, RNA lysates were obtained at different timepoints during the differentiation and these were subsequently analysed for relevant marker genes by RT-qPCR.

As detailed in Figure 3.10 A, both *Oct4* and *Nanog* expression decreases rapidly with the onset of differentiation with no significant differences between GNS G treated cells and the DMSO control. As transient marker of the differentiation process *Brachyury* and *GATA5* were investigated (Figure 3.10 B). Brachyury gene expression peaks early in differentiation around day 1. When comparing GNS G and DMSO treated cells, it becomes clear that there is a slight increase in *Brachyury* expression in GNS G treated cells on day 1. However, the overall expression pattern of the marker is maintained. For *GATA5*, there are no significant differences between cells with elevated O-GlcNAc levels and the control. As late markers of this differentiation, cardiac troponin T (*cTNT*) and *HAND1* were chosen (Figure 3.10 C). Both markers are induced to a comparable degree in both GNS G treated cells and the control. However, slight differences can be observed at intermediate stages of the differentiation, i.e. for *HAND1* expression on day 7. Taken together these results demonstrate that ChiPS4 differentiation into the mesodermal lineage remains largely unaffected by increase O-GlcNAcylation in response to GNS G.

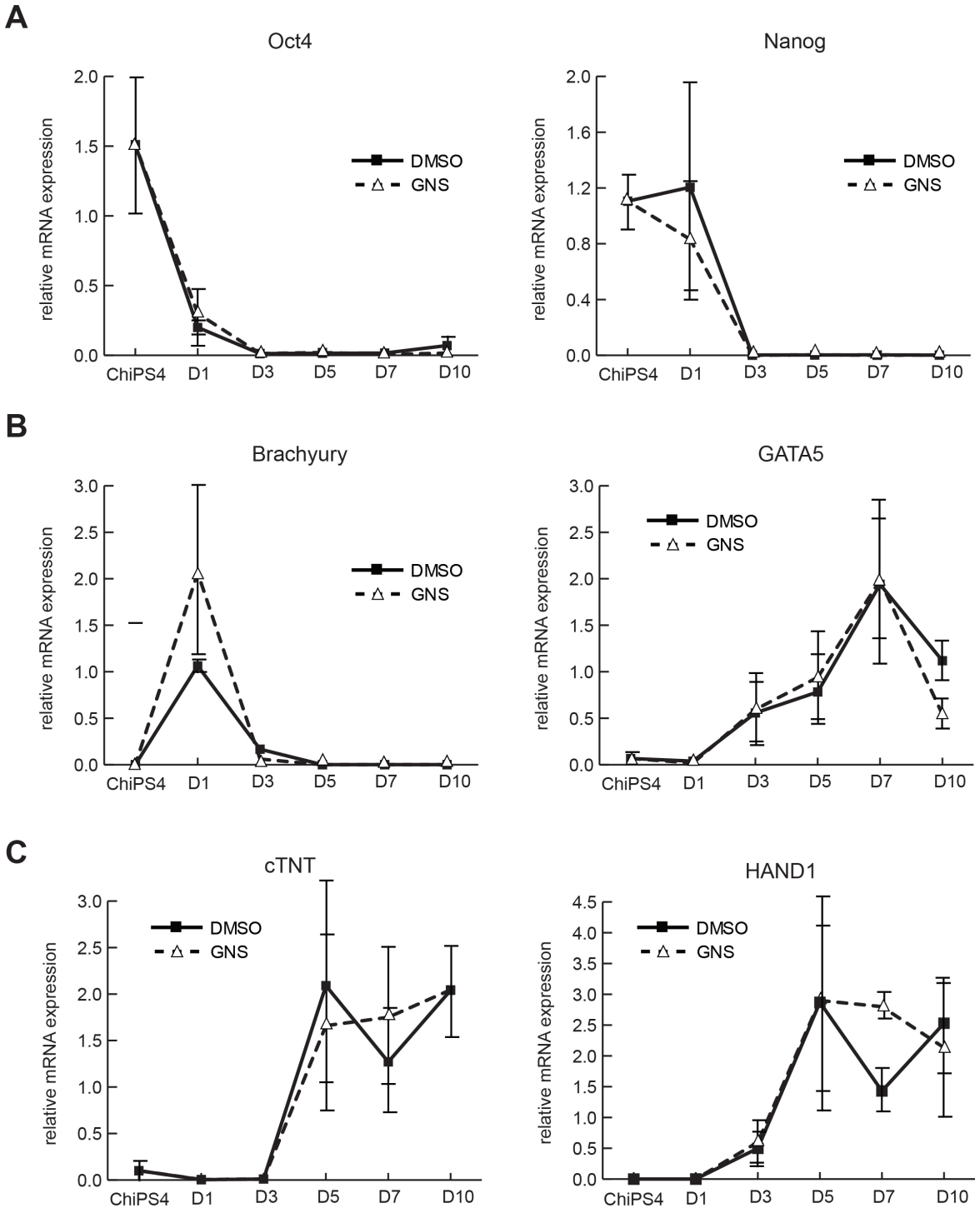

Figure (3.10) RT-qPCR analysis of marker gene expression during cardiomyocyte differentiation. A: mRNA levels of the pluripotency markers *Oct4* and *Nanog*; **B**: gene expression profiles of the transient differentiation markers *Brachyury* and *GATA5*; **C**: expression of the cardiomyocyte markers *cTNT* and *HAND1*. n=3 independent biological repeats, errorbars represent SEM.

In addition to the analysis of marker gene expression, global O-GlcNAcylation at different timepoints during cardiomyocyte differentiation was analysed by Western Blot. The result of this is displayed in Figure 3.11. It is clear that O-GlcNAcylation initially decreases with the onset of differentiation. Throughout the differentiation process, O-GlcNAc levels remain at a low level and only increase again at late stages of the differentiation (day 10). Overall the pattern of O-GlcNAcylation displays little dynamics. Interestingly, treatment with GNS G appears to not only raise total O-GlcNAcylation, but it also induces peaks of high O-GlcNAcylation at intermediate timepoints of differentiation (i.e. day 5). These findings might suggest a role of O-GlcNAc that modifies intermediate differentiation processes in this assay. However, cells appear to be able to commit to the mesodermal lineage regardless of those modifications.

3.3.3. O-GlcNAc functions in ectodermal differentiation of ChiPS4

GNS treatment hinders neural differentiation of ChiPS4

Previous experiments in mouse ES cells had shown a pronounced effect of GNS treatment on neural differentiation (Speakman et al., 2014). Therefore, a neural differentiation assay was chosen to test for effects of GNS on differentiation into ectodermal lineage. The differentiation was performed according to the well-established protocol published by Chambers et al. (2009). The protocol is based on high density culture of human iPS cells which are subjected to dual inhibition of SMAD signalling in order to direct their differentiation to a neural fate (Chambers et al., 2009). In accordance with the previously described experiments on endodermal and mesodermal differentiation, ChiPS4 cells were treated with GNS or DMSO and the expression of marker

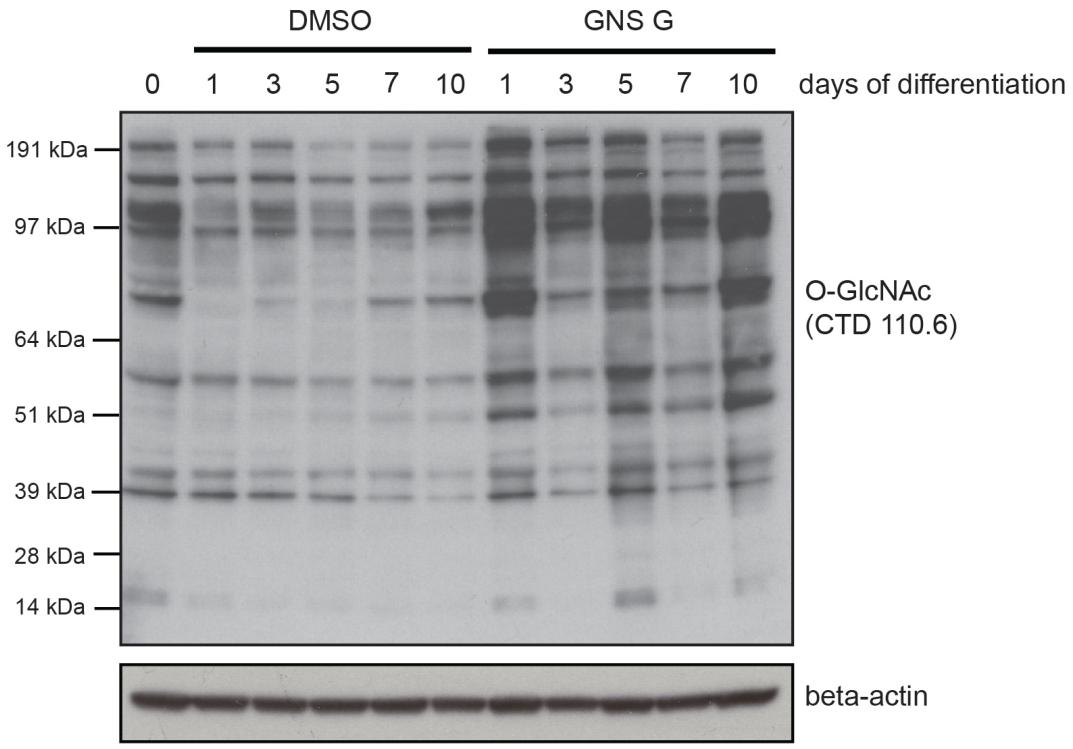

Figure (3.11) Variations in global O-GlcNAc levels during cardiomyocyte differentiation of ChiPS4. Analysis of O-GlcNAcylation on whole cell lysates of ChiPS4 treated with GNS G or DMSO at various timepoints during cardiomyocyte differentiation. A representative Western Blot is shown.

genes at various timepoints over the course of differentiation were examined. Initially, RT-qPCR analysis of the pluripotency genes *Oct4* and *Nanog* as well as the ectoderm markers *Pax6*, *Sox1*, *FoxG1* and *Pax7* was performed.

As can be seen from Figure 3.12 A, both *Oct4* and *Nanog* gene expression levels decrease rapidly with advancing differentiation. This quick drop in the expression of pluripotency markers is consistent throughout all the differentiation assays performed and no differences can be noted between cells that were treated with the vehicle control and GNS-treated cells.

Strikingly, dramatic differences between GNS treated cells and controls can be observed with regard to the expression of neural markers (Figure 3.12 B). Expression of *Sox1*, *FoxG1* and *Pax7* is strongly induced as a consequence of the cell's neural differentiation for DMSO treated cells. In sharp contrast, GNS treated cells do not exhibit expression of these neural marker genes. For GNS treated cells, *Sox1* and *FoxG1* gene expression remains at a basal level throughout the differentiation and is not induced at any stage. *Pax7* mRNA levels increase initially (until day five of the assay) and experience a sharp drop afterwards in GNS. Interestingly, *Pax6* expression does not appear to be affected by GNS treatment and is induced equally for control and GNS treated cells. These findings implicate that while GNS treated ChiPS4 loose their pluripotency, they are unable to commit to a neural fate as respective marker genes are mostly not expressed even at late stages of neural differentiation.

O-GlcNAcylation during neural differentiation

Analysis of global O-GlcNAcylation over the course of neural differentiation was again performed by Western Blot (Figure 3.13). It can be observed that O-GlcNAcylation during neural differentiation appears to be very dynamic. Overall levels of the mod-

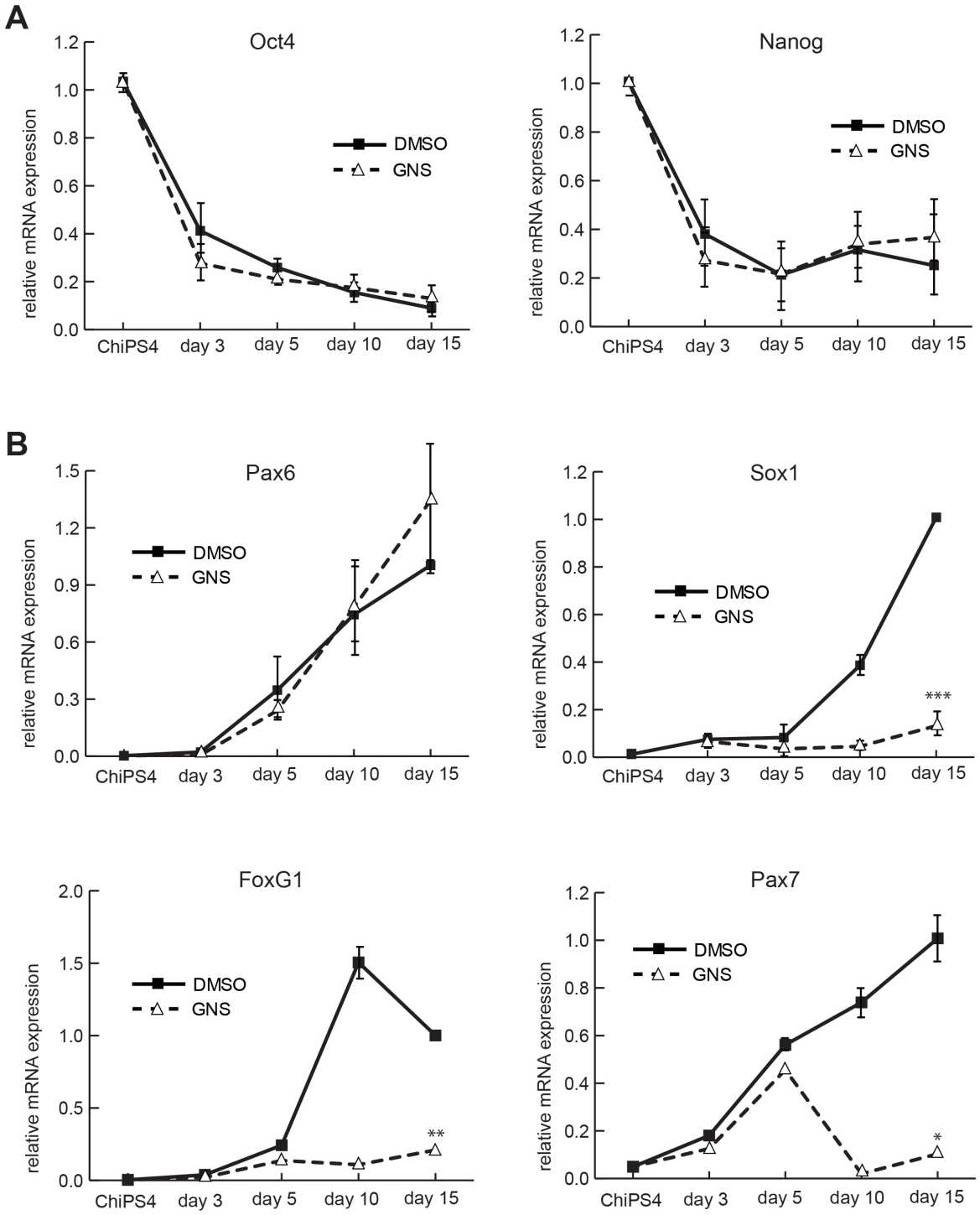

Figure (3.12) RT-qPCR analysis of marker gene expression during ectoderm differentiation. A: mRNA levels of the pluripotency markers *Oct4* and *Nanog*; **B**: gene expression profiles of the ectoderm markers *Pax6*, *Sox1*, *FoxG1* and *Pax7*. n=4 independent biological repeats for *Nanog*, *Oct4* and *FoxG1*; n=3 independent biological repeats for *Pax6*, *Sox1* and *Pax7*; for *FoxG1* and *Pax7* weighted means are shown. ***: $p \leq 0.001$; **: $p \leq 0.01$; *: $p \leq 0.05$ (Student's t-test, unpaired); errorbars represent SEM

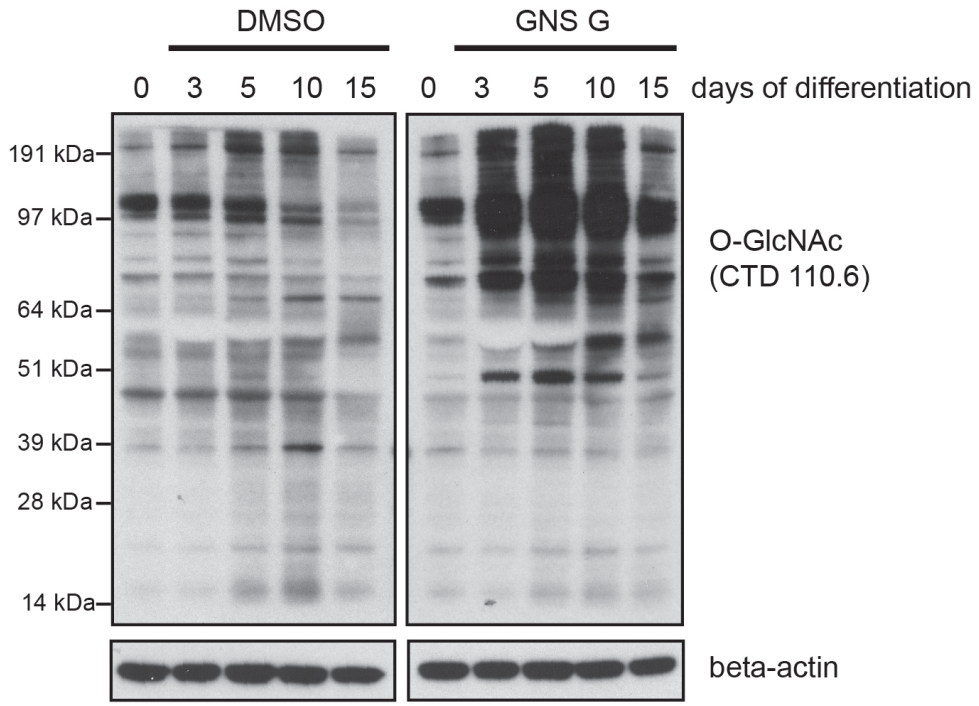

Figure (3.13) Variations in global O-GlcNAc levels during neural differentiation of ChiPS4. Left panel: O-GlcNAcylation of ChiPS4 treated with DMSO. Right panel: O-GlcNAcylation of GNS treated cells. A representative Western Blot is shown.

ification initially increase until day five. O-GlcNAc levels then decrease again and at day 15 of differentiation the level of the modification has returned to a similar level as before the start of differentiation. In addition, the pattern of O-GlcNAcylation varies significantly with a number of bands appearing and disappearing. For GNS treated cells the pattern appears to remain comparable to the DMSO controls and O-GlcNAc levels are respectively elevated in response to the drug treatment. It can be concluded from this analysis that during neural differentiation both O-GlcNAc levels and the pattern of O-GlcNAcylation are very dynamic underlining the importance of the modification for neural differentiation.

GNS hinders neural differentiation in embryoid bodies

The previously described observations were based on a high-density differentiation protocol. Therefore it is not possible to assess morphological abnormalities that might occur in GNS treated cells that can not complete neural differentiation. With the aim of circumventing this issue an additional neural differentiation assay was chosen which allows a close observation of cell morphology. The assay is based on initial differentiation of ChiPS4 in embryoid bodies (EBs) followed by plating at a stage at which cells should have differentiated to neuroectoderm (see Figure 3.14). Importantly, cells were treated with GNS at from two different stages of differentiation (day one and day nine) in order to determine at which stage abnormalities occur. As illustrated in Figure 3.14 B, cells of the vehicle control have generated distinct neural structures at day 15 of the assay. This is characterized by the honeycomb-like structure of neural rosettes. A very similar morphology can be observed for cells that had been treated with GNS only from day nine on (GNS D9). These cells as well exhibit a characteristic neural morphology. In contrast, cells that had been treated with GNS from the start

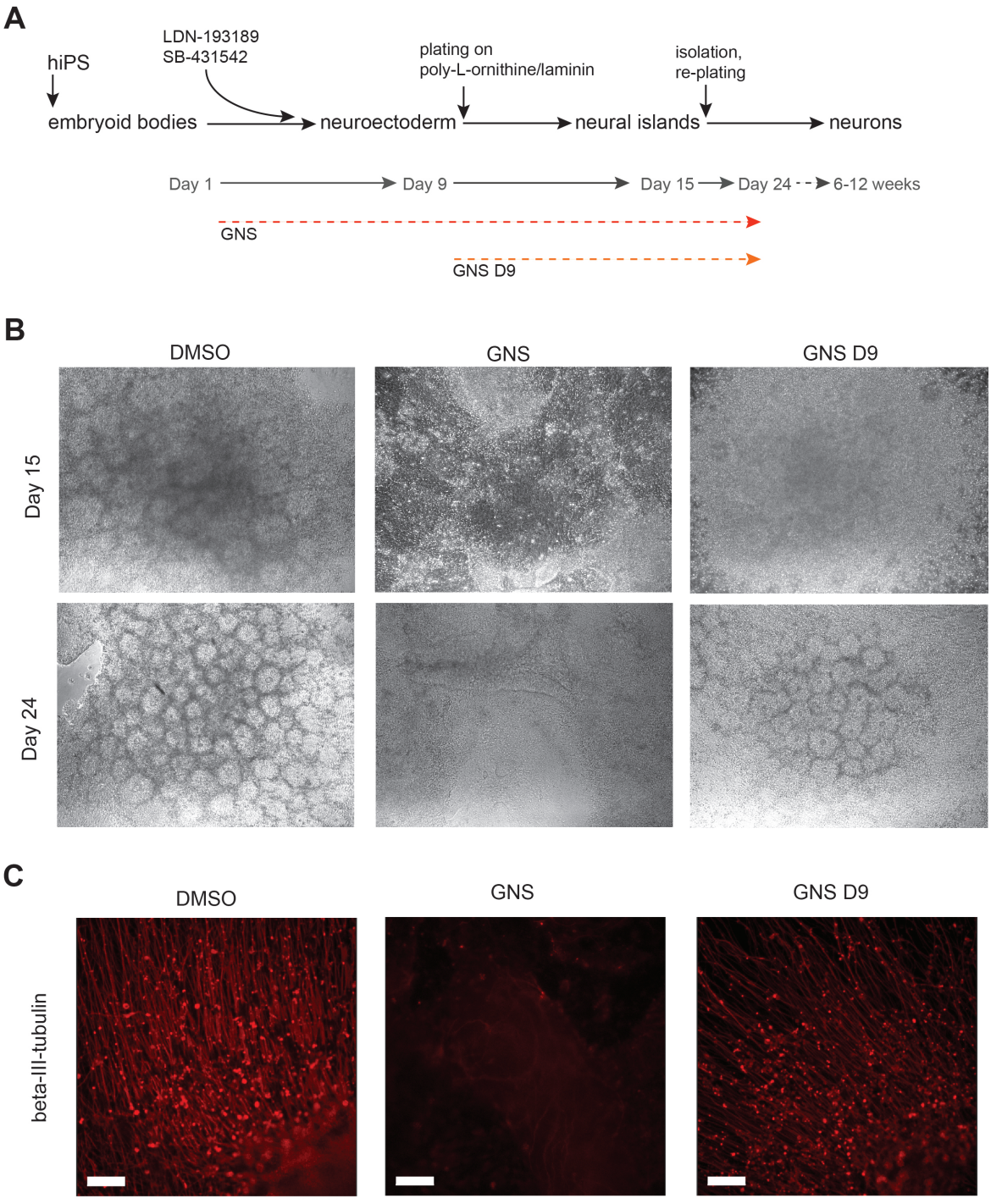

Figure (3.14) GNS G hinders neural differentiation in embryoid bodies. **A**: Schematic of the differentiation protocol for neural differentiation of ChiPS4 in embryoid bodies. **B**: Brightfield images of ChiPS4 at day 15 and 24 of the differentiation assay. **C**: Immunofluorescence analysis of the neuron marker β-III-tubulin at day 24. Representative images are shown. Scalebar = 90 μm

of differentiation do not show any organized morphology and appear to be randomly clustered together.

This phenotype is even more pronounced at day 24 as cells have flattened out further and structures are less compact: Both the DMSO control and GNS D9 cells demonstrate a very organized morphology and between neural tube-like structures cells have started to form mature and polarised neural rosettes. For cells that had been treated with GNS throughout the assay, it can also be observed that the cell layer has become flatter. However, cells remain largely unorganized and completely lack neural morphology.

In addition, the ability of the cells to generate neurons was assessed. Generally, neurons form in the periphery of EBs in this assay. To determine neuron formation ability, cells were fixed, immunostained and analysed for the neuron marker β-III-tubulin. As illustrated in Figure 3.14 C, DMSO treated control cells generate a large number of neurons that extend from the original EB (left image). The same observation can be made for cells that had only had GNS from day nine of the differentiation assay, although they appear to generate slightly fewer neurons (right image). When comparing to cells that had been treated with GNS from the beginning of the assay, it is clear that those cells are largely unable to differentiate into β-III-tubulin-positive neurons.

These observations underline the previously described findings and emphasize that GNS treatment inhibits neural differentiation. Importantly, if GNS is added at a later stage of differentiation cells progress normally. This implies that early events of neural differentiation are strongly affected by GNS, but that there is no impact at later stages of differentiation.

GNS treatment disrupts neural rosette formation

During neural differentiation, human iPS cells show certain morphological characteristics, i.e. the formation of neural rosettes. These are two-dimensional structures of neural progenitors in which the cells exhibit strong polarity towards the centre of the rosette. More mature neurons are then usually found to expand from the central point.

With the aim of analysing rosette formation ability, cells were differentiated according to the EB differentiation protocol above. At day 20 cells were fixed and stained for different markers. Staining for β-III-tubulin and Sox2 was performed in order to analyse for neuron formation as well as for commitment to a neural progenitor fate. Representative images of the analysis are depicted in Figure 3.15. Figure 3.15 A illustrated that ChiPS4 of the DMSO control form distinct rosettes. As can be seen from the β-III-tubulin staining (A') several neuron extend from the centre of the rosette. Most cells are Sox2-positive (A") confirming their neural progenitor identity. When comparing to cells that were treated with GNS from day nine onwards (Figure 3.15 B), it is clear that these cells are also organized in rosettes. However, fewer neurons can be found and Sox2 expression appears to be slightly weaker than in the DMSO control. Strikingly, ChiPS4 that were treated with GNS throughout the differentiation did not show any organized structure at all (Figure 3.15 C). In addition, these cells are also not able to generate mature neurons. Even commitment to a neural fate appears to be severely obstructed as most cells are negative for Sox2.

Neural rosette formation can also be evaluated by investigating cell polarity. Therefore, staining was performed for ZO-1 and N-Cadherin. Both proteins mediate cell polarity in the centre of rosettes. From Figure 3.16 A it becomes apparent that both ZO-1 and N-Cadherin are localized in the centre of neural rosettes in DMSO treated

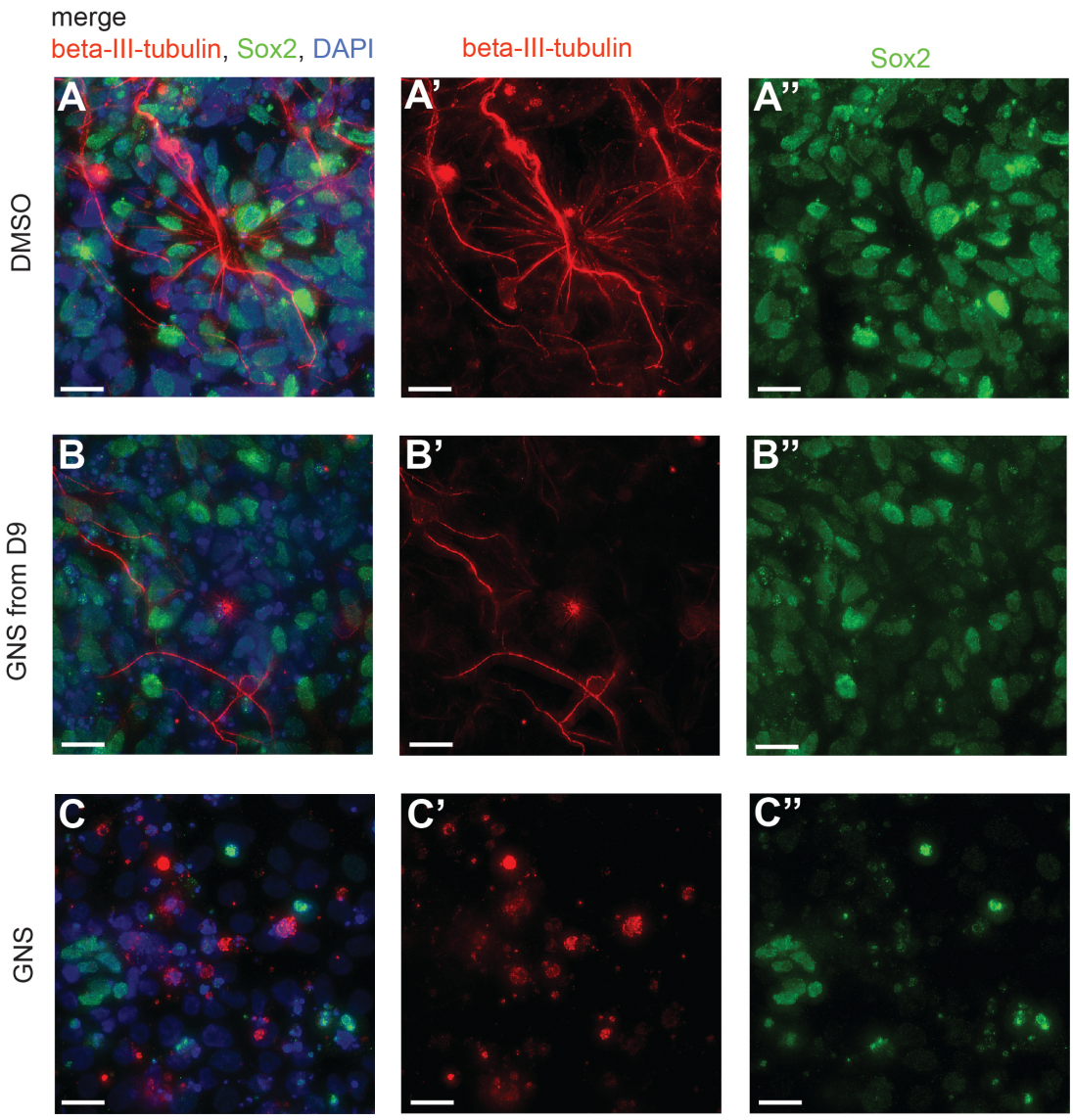

Figure (3.15) Immunofluorescence analysis of ChiPS4 after neural differentiation in embryoid bodies. Immunofluorescence staining for β-III-tubulin and Sox2 demonstrates that neural differentiation is hindered by GNS (**C**) whereas control cells (**A**) or cell treated at a later timepoint (**B**) differentiate normally. Representative images are shown. Scalebar = 20 μm

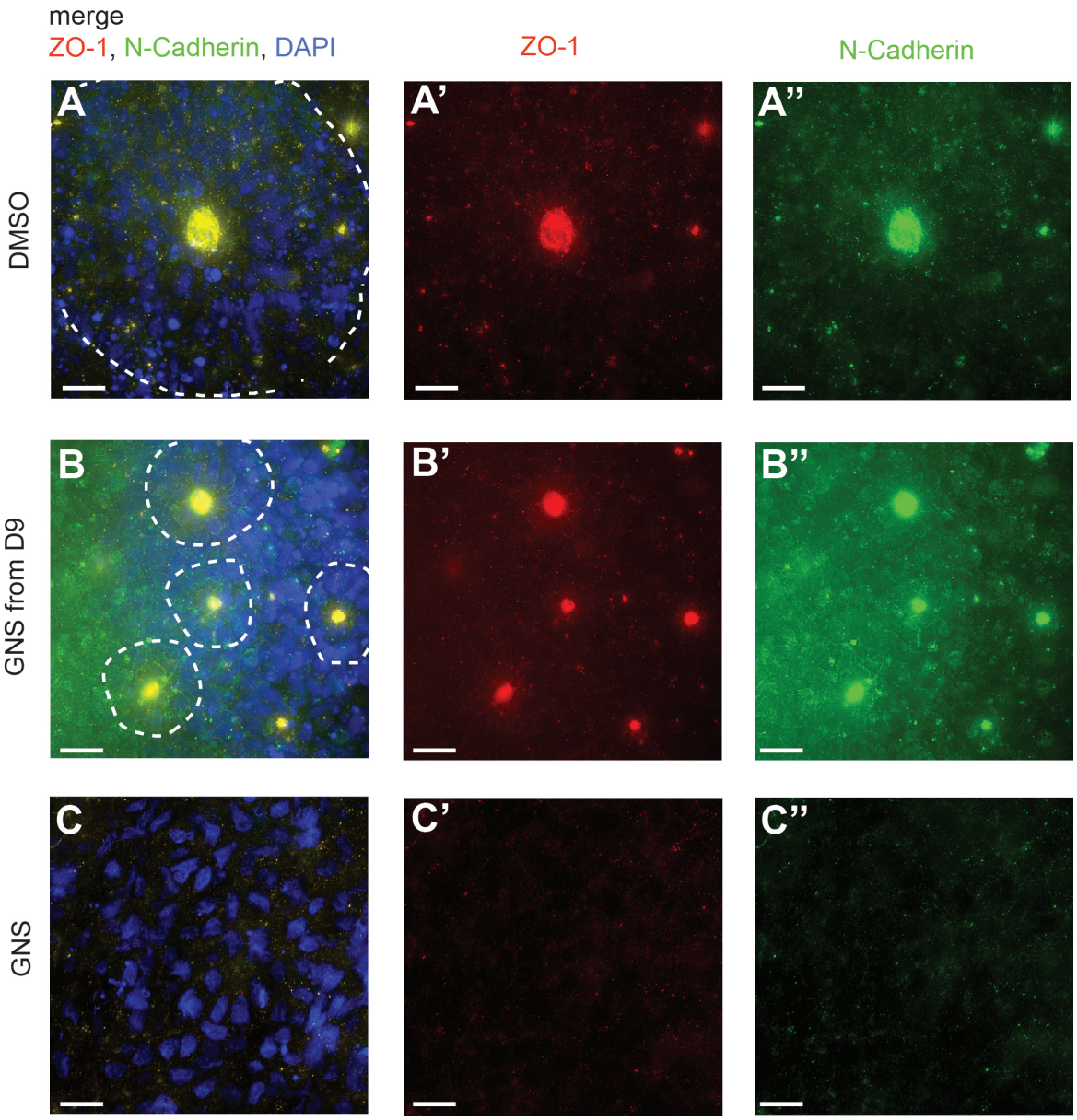

Figure (3.16) Immunofluorescence analysis of ChiPS4 after neural differentiation in embryoid bodies. Immunofluorescence staining for ZO-1 and N-Cadherin indicates that polarity of neural rosettes is disrupted by GNS treatment (**C**) in comparison to control cells (**A**) or cell treated at a later timepoint (**B**). Representative images are shown. Scalebar = 20 μm

ChiPS4 demonstrating central polarity of the cells. The same is true for GNS D9 cells which also exhibit cell polarity to the rosette's centre as shown by the co-expression of ZO-1 and N-Cadherin (Figure 3.16 B). In contrast, GNS treated cells are devoid of any central polarity and neither ZO-1 nor N-Cadherin are expressed to a high level (Figure 3.16 C). These findings underline the previously presented observations with regard to the strong impact of GNS in ectodermal differentiation.

GNS addition at early differentiation stages impairs ectoderm differentiation and neurogenesis

As the previously described results demonstrate that high O-GlcNAc levels impair ectodermal differentiation, further experiments were performed to investigate if these effects are restricted to a certain stage of differentiation.

Monolayer neural differentiation was performed, in order to determine if ectoderm differentiation was still hindered if GNS G was added at later stages. ChiPS4 were differentiated for ten days and GNS G was added at different timepoints. Cells were then harvested for RNA and RT-qPCR analysis for the marker Sox1 was performed. Figure 3.17 illustrates the result of this analysis. It can be observed that Sox1 expression is higher the later during differentiation GNS G was added to the cells. Cells that had only been treated with GNS G from day 3 of the differentiation onwards express Sox1 to an extent comparable with cell that had been treated with a DMSO control. This finding demonstrated clearly that increased O-GlcNAcylation only impairs ectoderm differentiation if it occurs during early stages.

With the aim of validating these data, neural differentiation of ChiPS4 in EBs was performed and GNS was again added at different timepoints. In contrast to the previously described method, cells were treated with BDNF and GDNF at the later

stages of differentiation (day 15 onwards) to induce neurogenesis. At day 24, cells were fixed and analysed by immunofluorescence staining for the neuron marker β-III-tubulin. These results are depicted in Figure 3.17. Cells that were treated with GNS G before day 3 exhibit a massively decreased ability of neuron formation. However, if GNS G had been added after day 2, neurogenesis is no longer impaired and the neuron formation of those cells is comparable to the DMSO control.

Taken together this findings illustrate that high levels of O-GlcNAc only impair ectodermal differentiation if they persist at early stages of differentiation. Addition of GNS G after the initial stages of commitment to this lineage have no effect on differentiation and neurogenesis.

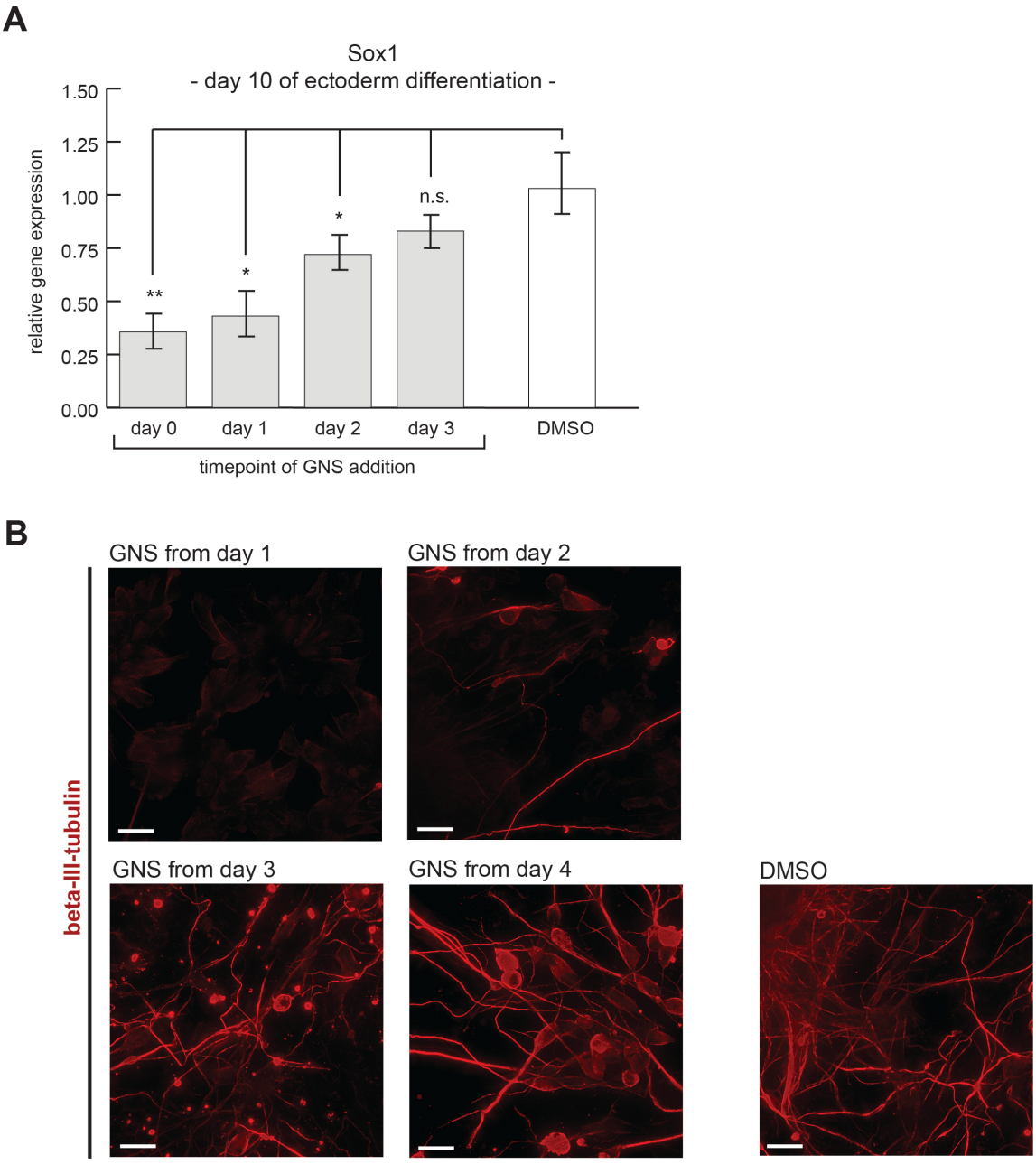

Figure (3.17) GNS G treatment impairs ectodermal differentiation of ChiPS4 only if added at initial stages. **A**: RT-qPCR analysis of the ectoderm marker gene *Sox1* (n=3 independent biological repeats, errorbars represent SEM; **: $p \leq 0.01$; *: $p \leq 0.05$; n.s.: $p > 0.05$ (Student's t-test, unpaired)). **A** Immunofluorescence staining for β-III-tubulin on ectoderm differentiation in EBs indicates that neurogenesis is disrupted if GNS G is added before day 3 of differentiation. Representative images are shown. Scalebar = 20 μm

4. The role of O-GlcNAc in transcriptional regulation

The previously described findings highlight the importance of O-GlcNAc in pluripotent stem cell differentiation. Differentiation of ES or iPS cells is a tightly regulated process which is mediated by changes in the epigenetic landscape and gene expression. In order for a cell to make the transition from a pluripotent cell to a mature somatic cell it is vital that some genes are increasingly expressed while other are downregulated or entirely repressed. With the aim of deciphering the effects of O-GlcNAcylation on gene expression, microarray analysis was performed on 46C mouse ES cells treated with GNS C or DMSO as a vehicle control (Speakman et al., 2014). Gene set enrichment analysis (GSEA) on this data indicated that two gene sets that are typically silenced by epigenetic mechanisms are reexpressed in GNS treated cells. These two gene sets are the Polycomb repressive complex repressed genes (PRCR) and the two-cell (2C) stage gene set which is characteristic for the early developmental stage of the two-cell embryo. The following experiments were performed with the aim of validating and extending the microarray data presented in Speakman et al. (2014).

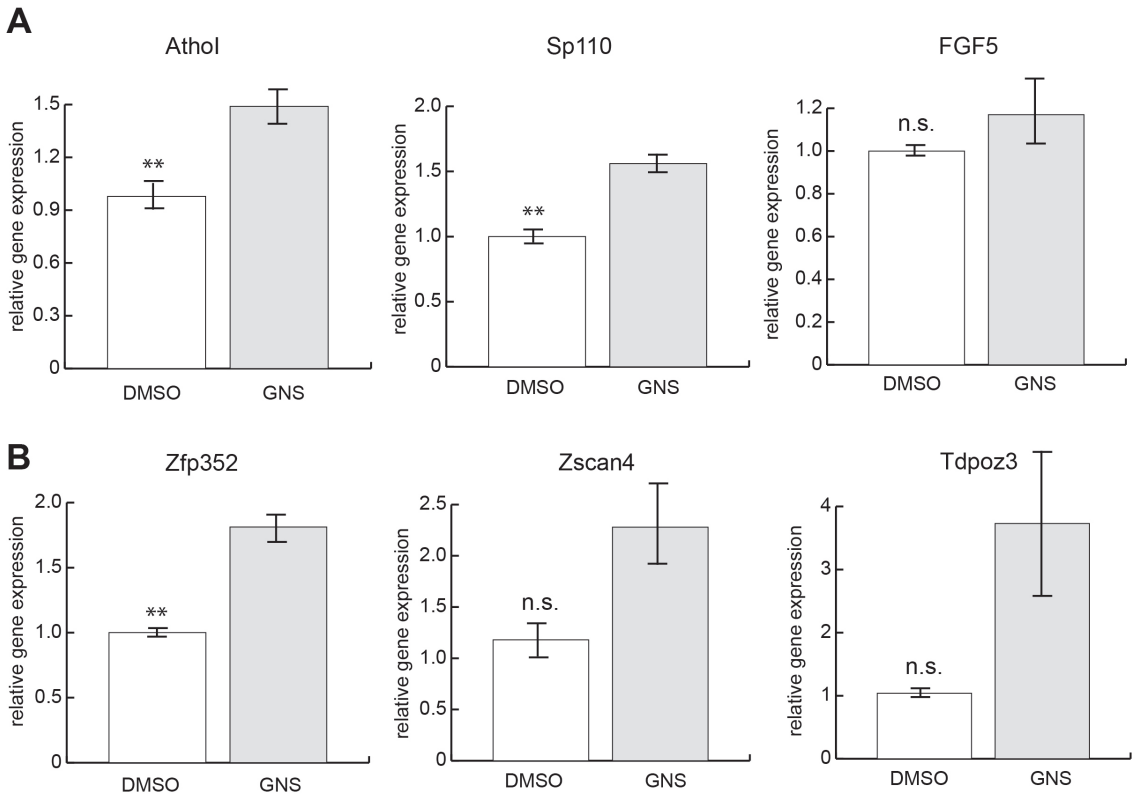

Figure (4.1) Polycomb repressive complex repressed genes and 2C genes are upregulated by GNS treatment. **A**: RT-qPCR analysis of selected genes of the PRCR geneset, n=6 independent biological repeats. **B**: RT-qPCR analysis of selected 2C genes; n=3. ***: $p \leq 0.001$; **: $p \leq 0.01$; *: $p \leq 0.05$; n.s.: $p > 0.05$ (Student's t-test, paired); errorbars represent SEM.

4.1. Oga inhibition or knockdown causes reexpression of epigenetically silenced genes

In order to validate the findings of the GSEA, the expression of selected candidate genes from the 2C and PRCR geneset in cells with high O-GlcNAc levels was determined. For this, E14 cells were treated with GNS G or the respective vehicle for two days and subjected to RT-qPCR analysis. The findings of these experiments are depicted in Figure 4.1. With a view to the expression of genes that are repressed by polycomb repressive complexes, mRNA levels of *AthoI*, *Sp110* and *FGF5* were analysed (Figure 4.1 A). It is clear from these data that expression of *AthoI* and *Sp110* is significantly increased in GNS treated cells when compared to the DMSO control. However, FGF5 expression appears to be more variable with only a small trend towards upregulation in GNS treated cells. *Zfp352*, *Zscan4* and *Tdpoz3* have been described as characteristic for the 2C subpopulation. All three genes experience a significant increase in their mRNA levels in cells with elevated O-GlcNAcylation in response to GNS treatment. Again it remains to be noted that the described effect appears to be more variable for some genes (e.g. *Tdpoz3*). In summary, the RT-qPCR data is in good agreement with the previous observations based on GSEA and it can be concluded that the expression of selected 2C and PRCR genes is upregulated by GNS treatment.

4.2. Knockdown of Oga leads to upregulation of 2C genes

With the aim of substantiating the previously described findings, knockdown experiments with siRNAs were applied in combination with gene expression analysis of 2C

genes. Using pools of siRNAs targeting Oga or Ogt it was possible to dramatically decrease protein levels of the respective protein. Figure 4.2 A illustrates the effect of siRNA mediated knockdown over a period of four days. It can be observed that a successful knockdown of Oga or Ogt can be achieved for as long as 72 hours. However, this effect is partially lost approximately 96 hours post-transfection. Importantly, in both cases a direct feedback downregulation of the antagonizing enzyme can be observed, i.e. knockdown of Oga leads to a feedback decrease in Ogt protein levels and vice versa. In addition the effects on mRNA levels of Oga/Ogt under knockdown conditions were investigated (Figure 4.2 B and C). Levels of Oga mRNA decrease rapidly in response to both Oga and Ogt knockdown. For Ogt, mRNA levels drop sharply after siOgt transfection (Figure 4.2C). Interestingly, after transfection with siOgt a rise in Ogt mRNA levels could be observed. This might hint at a compensatory transcriptional effect in response to the feedback regulation that can be observed at protein level.

Following the validation of the siRNA mediated knockdown for both enzymes, expression of 2C genes after a 48 h knockdown of Oga was monitored. As illustrated in Figure 4.2 D cells transfected with siOga exhibit an increased expression of the three selected 2C genes in comparison to cells transfected with a non-targeting control. However, an increase in variability between independent experiment was observed compared to the analysis that had been performed on GNS treated cells (compare Figure 4.1 B). These findings support the previous observations and demonstrate that both Oga inhibition and knockdown lead to a re-expression of genes characteristic for the 2C subpopulation.

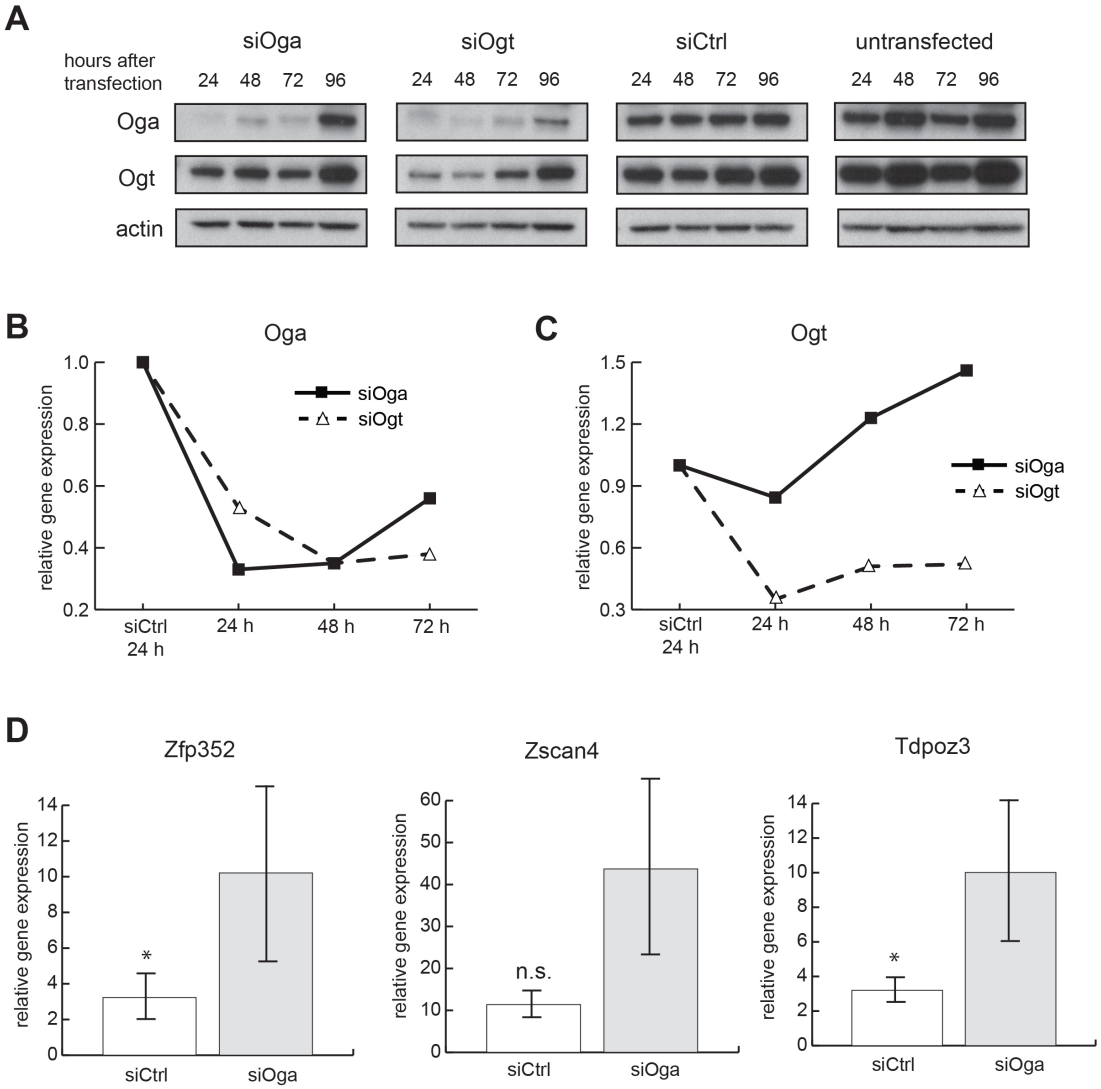

Figure (4.2) Knockdown of Oga using siRNAs induces re-expression of 2C genes. **A**: Oga and Ogt protein levels at different timepoints after siRNA transfection show a decrease in respective protein levels. A representative Western Blot is shown. **B, C**: mRNA levels of Oga and Ogt in response to siOga/siOgt; n=2 independent biological repeats, mean values are shown. **D**: Gene expression of 2C genes is elevated in siOga; n=3 biological repeats; *: $p \leq 0.05$, n.s.: $p > 0.05$ (Student's t-test, paired); errorbars represent SEM.

4.3. GNS G treatment induces 2C promoter activity

With the aim of further validating the previously described results on the 2C subpopulation, the activity of the 2C promoter in response to GNS G treatment was investigated. For this purpose, two MERVL (Mouse endogenous retrovirus with a leucine tRNA primer binding site) reporter cells were generated (cell lines generated by Marios Stavridis). To generate the reporter, the 5'LTR of the MERVL retrovirus-like element, the primer binding site and a portion of the Gag gene was cloned upstream of EGFP (enhanced green fluorescent protein) or RFP (red fluorescent protein) as described by Macfarlan et al. (2012) and stable E14 mouse ES cell lines expressing the respective reporter were generated.

Cells were treated for 48 hours with GNS G, DMSO as a negative control or the histone deacetylase inhibitor trichostatin A (TSA) (treatment for 2-4 hours) as a positive control and were then analysed by flow cytometry. The respective results are depicted in Figure 4.3 and Figure 4.4. Both reporter lines indicate a small but significant increase in 2C promoter activity. It is also important to note that the induction of the 2C promoter by GNS G appears to be comparable to treatment with TSA. Interestingly the number of reporter-expressing cells varied between the two reporter cell lines with a higher percentage of cells that were positive for the reporter in the 2C:EGFP line ($\sim$ 19 % EGFP-positive cells in DMSO) than in the 2C:RFP line ($\sim$ 11 % RFP-positive cells in DMSO). Regardless of these differences between the two lines the result remained consistent and clearly supports the previously presented data by indicating that GNS G treatments leads to an increase in 2C promotor activity.

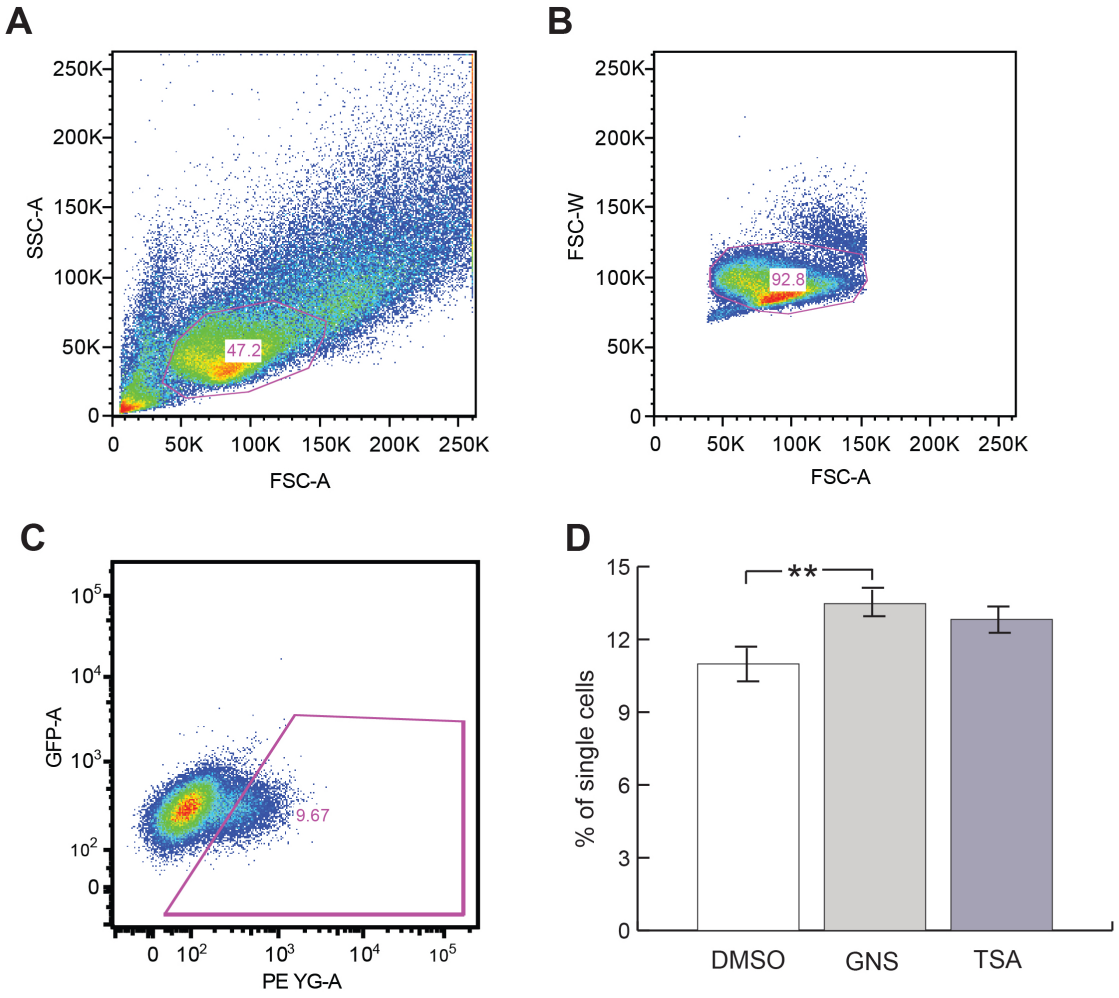

Figure (4.3) GNS G induces 2C promoter activity in E14 2C:RFP reporter cells. A: Gating strategy to exclude debris. B: Gating to exclude duplets. C: Gating to select RFP-positive cells while excluding autofluorescent cells. D: Percentage of RFP^+ single cells. n=4 independent biological repeats, **: $p \leq 0.001$ (Student's t-test, paired). Errorbars represent SEM.

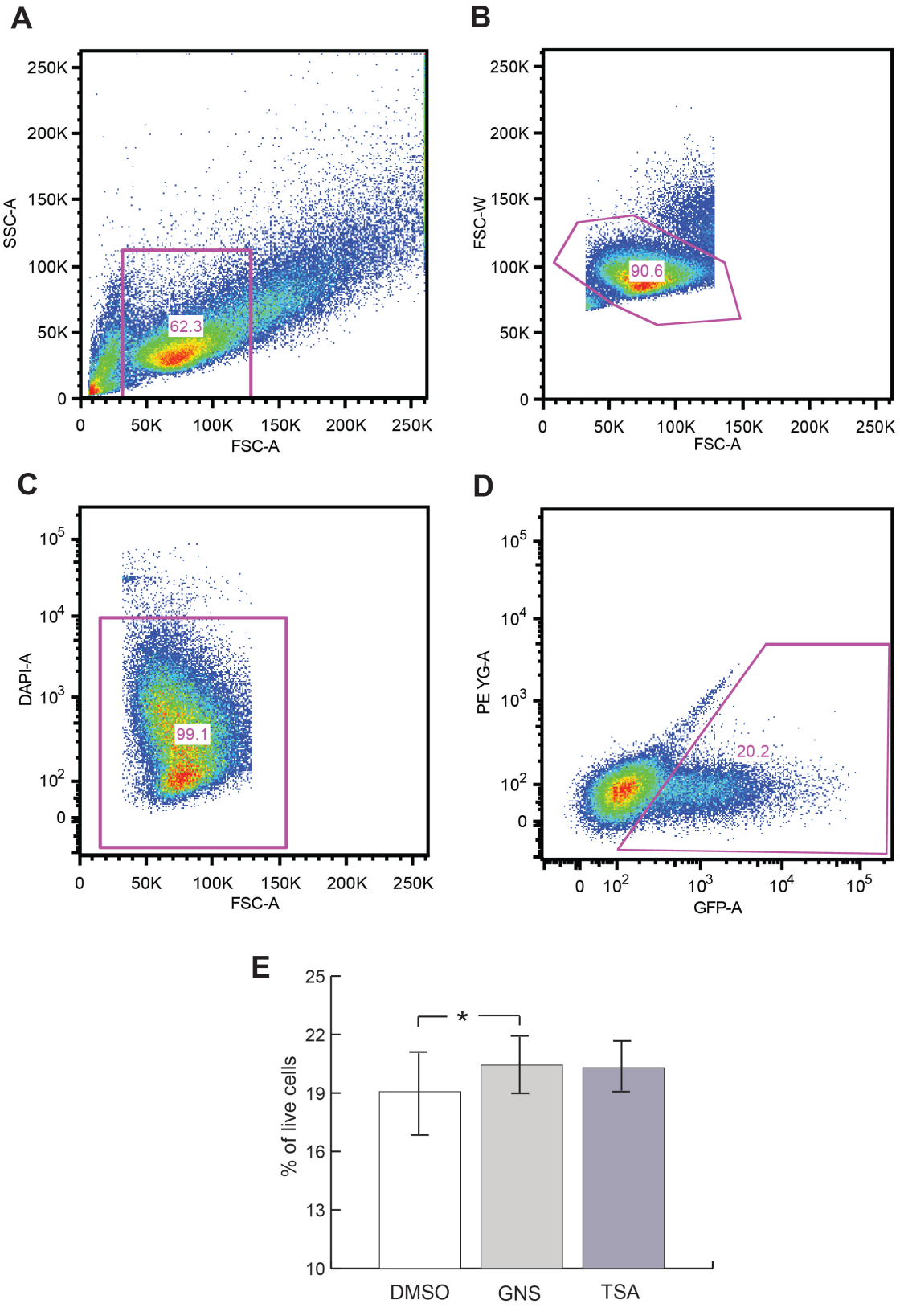

Figure (4.4) GNS G induces 2C promoter activity in E14 2C:EGFP reporter cells. A: Gating strategy to exclude debris. B: Gating to exclude duplets. C: Gating strategy to exclude dead cells (= DAPI-positive). D: Gating to select EGFP-positive cells while excluding autofluorescent cells. D: Percentage of EGFP$^+$ live cells. n=3 independent biological repeats, *: $p \leq 0.001$ (Student's t-test, paired), errorbars represent SEM.

4.4. Effects of GNS G on total histone acetylation

The data above suggested that effects of GNS G on mouse ES cells might to some extend be comparable to TSA treatment. TSA is a broad inhibitor of histone deacetlyases and therefore increases histone acetylation in a global manner in the cells. Elevated histone acetylation leads to an open chromatin conformation and therefore to an increase in transcriptional activity. It remains to be clarified to which degree the re-expression of 2C and PRCR gene sets by GNS G could be the results of a global mechanism such as histone acetylation. Previous studies suggest that the C-terminus of Oga contains a domain which can exhibit histone acetyltransferase (HAT) activity (Toleman et al., 2004). As GNS G treatment results in a feedback upregulation of Oga (section 3.1.2), further investigations aimed to clarify whether this would also correlate with an increase in histone acetylation and therefore global transcriptional activity.

Initially, histone extracts were generated from E14 mouse ES cells treated with GNS G for 48 hours. These extracts were then analysed by Western Blot using an antibody against acetylated lysine. As illustrated in Figure 4.5 A a strong increase in lysine acetylation of histone 3 can be observed. However, it needs to be noted that those results were highly variable between different experiments.

In order to directly relate those findings to Oga protein levels, knockdown and overexpression of Oga was performed and histone extracts were again analysed for levels of lysine acetylation. The results of these experiments are depicted in Figure 4.5 B. Upon knockdown of Oga or Ogt no significant differences in lysine acetylation could be observed when comparing to the respective control. In contrast, overexpression of Oga appears to lead to a small increase in lysine acetylation of histone 3, but not histone 4. Again, results from these experiments were highly variable with low

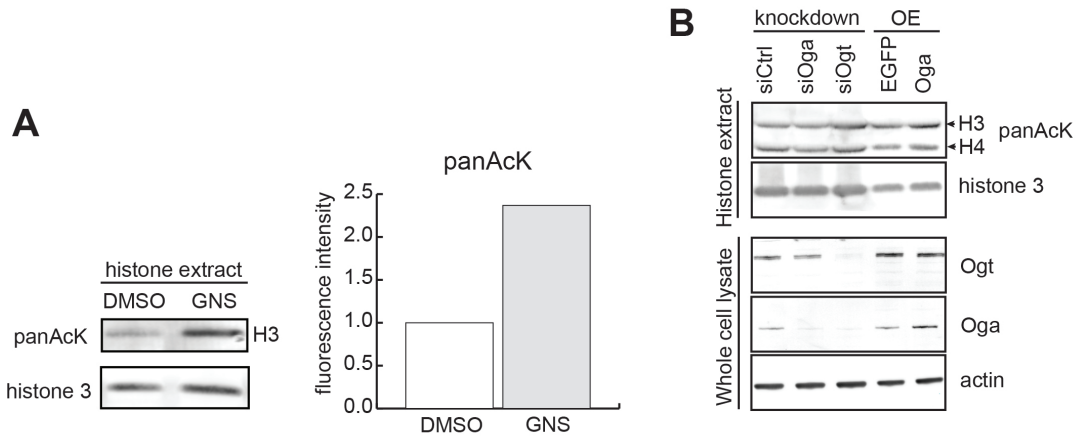

Figure (4.5) GNS G increases global histone 3 lysine acetylation. A: Western Blot analysis of total lysine acetylation of histone 3 on histone extracts after GNS G treatment (left panel). Right panel: Western Blot quantification. **B**: Western Blot analysis of panAcK on histone 3 on histone extracts after knockdown or overexpression of Oga. Both Western Blots were selected to demonstrate a possible effect of GNS on lysine acetylation.

reproducibility and therefore difficult to interpret. For these reasons it was decided to analyse histone acetylation by another method.

Therefore, global histone 3 acetylation was investigated by ELISA. E14 mouse ES cells were transfected with two Oga overexpression plasmids: Oga full-length (Oga FL) and a naturally occurring splicing variant (abbreviated Oga SV) which is devoid of the majority of the potential HAT domain. The respective results are illustrated in Figure 4.6. It is evident from the data obtained that there is a slight trend towards an increase in histone 3 acetylation in cells that overexpress full-length Oga. This effect is lost when the splicing variant that lacks the potential HAT domain is overexpressed. However, it is also clear that there is a very high variability between different experiments which make it difficult to interpret the result. These issues appear to be due to a technical problem that could be linked with batch effects of the ELISA kit that was used. Due to the high variability of both the Western Blot analysis and also the results obtained from the histone 3 acetylation ELISA it is unclear if GNS G treatment or

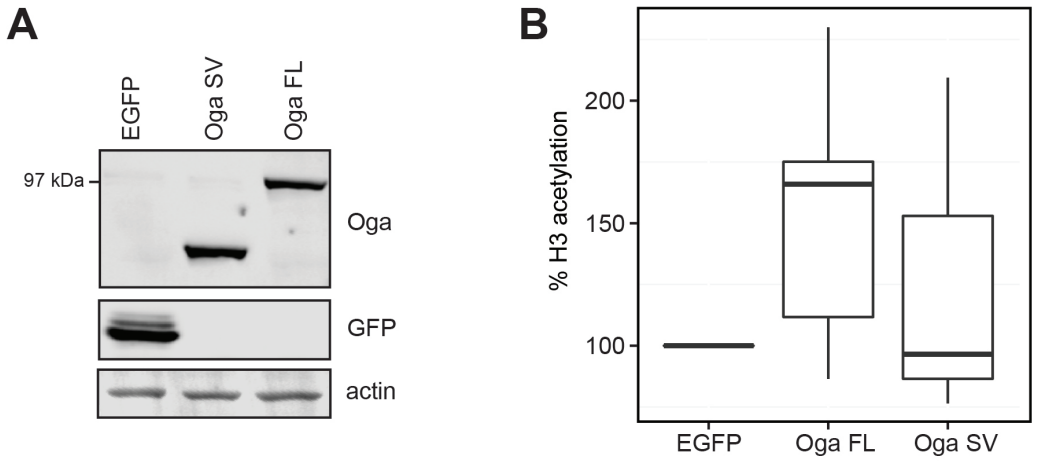

Figure (4.6) Analysis of histone 3 acetylation by ELISA. A: Overexpression of EGFP (negative control), Oga splicing variant devoid of the potential HAT domain (Oga SV) and full-length Oga (Oga FL). **B**: ELISA of histone 3 acetylation. n=5 independent biological repeats, p=0.124 (Student's t-test, paired).

Oga overexpression lead to an increase in histone acetylation. Overall the observed effects appear to be insignificant and are most likely not the main mechanism leading to the previously described re-expression of silenced genes.

4.5. Effects of GNS G on DNA methylation

A number of recent publications suggest an interplay of O-GlcNAcylation and its regulators with Tet proteins (Shi et al., 2013; Deplus et al., 2013; Chen et al., 2013; Ito et al., 2013). Tet proteins mediate changes in DNA methylation, an important epigenetic mark involved in gene silencing. For these reasons the DNA methylation status of cells under high O-GlcNAc conditions was investigated using multiple approaches and the respective results are presented in the following sections.

4.5.1. Effects of GNS G on selected DNA methylation sites

Initial investigations were focussed on determining potential effects of high O-GlcNAc on DNA methylation at specific sites of selected genes. For this purpose, bisulfite sequencing on DNA of E14 mouse ES cells was performed. This method allows detection of methylated DNA by conversion of cytosine to uracil. Methylated sites are protected from the conversion which can be detected during the following sequencing. The genes investigated were the housekeeping gene *beta-actin*, the pluripotency-related genes *Rex1* and *Klf4* as well as the silenced gene *Rest* and the long non-coding RNA *Xist*. The respective results are illustrated in Figure 4.7, Figure 4.8, Figure 4.9, Figure 4.11 and Figure 4.10.

For *beta-actin* it can be observed that the number of methylated sites is slightly increased in GNS treated cells. However, a proportion of sites were lost during the sequencing (depicted as 'not present') which impose limitations in interpreting these data.

For *Rex1* (Figure 4.8) and *Klf4* (Figure 4.9) no major differences in DNA methylation can be observed when comparing DNA of GNS G treated cells to DNA of the

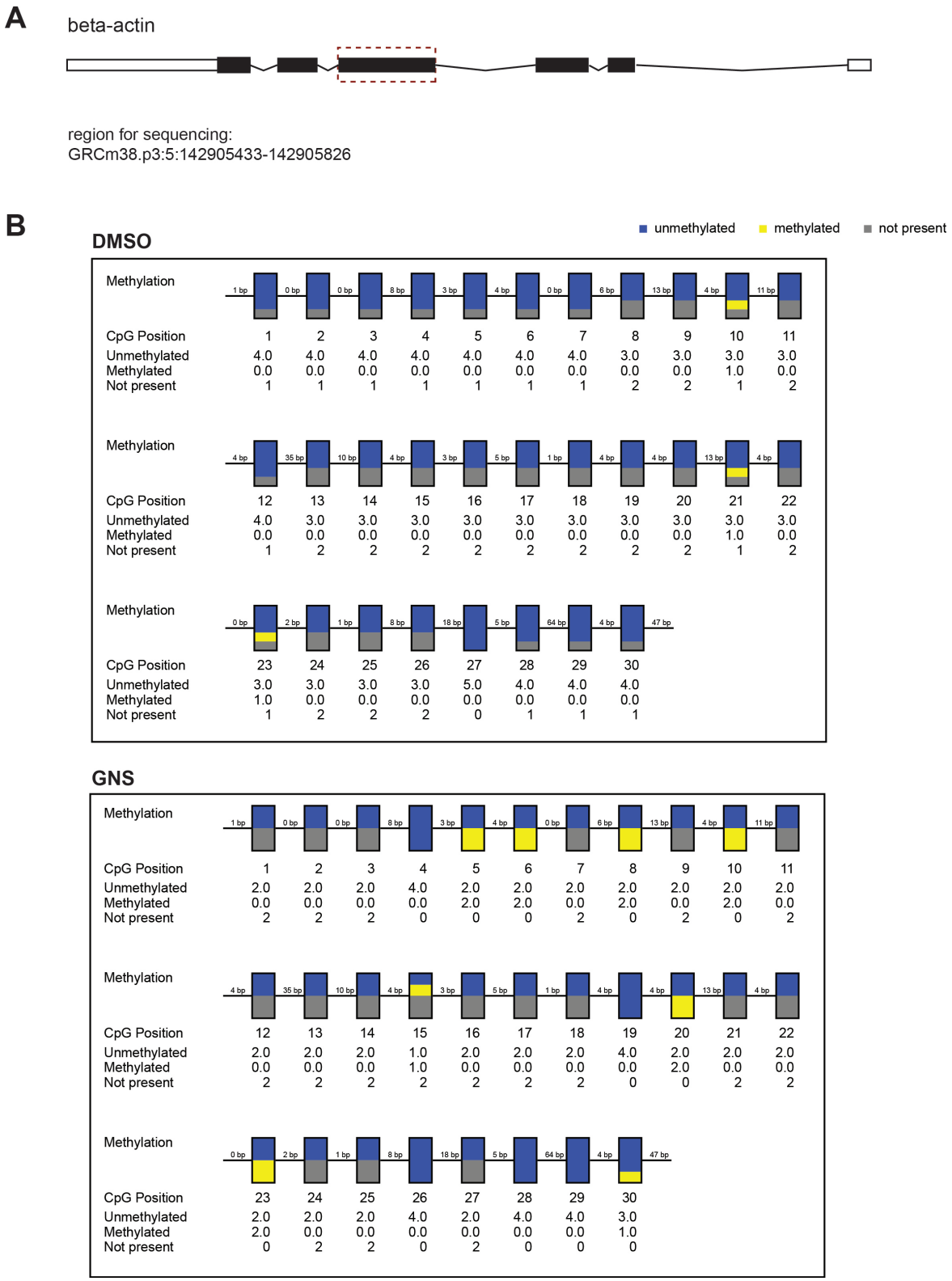

Figure (4.7) Bisulfite sequencing of beta-actin. A: Illustration of the region investigated by bisulfite sequencing. B: Methylation of CpG islands within the selected region in DMSO (upper panel) and GNS G (lower panel).

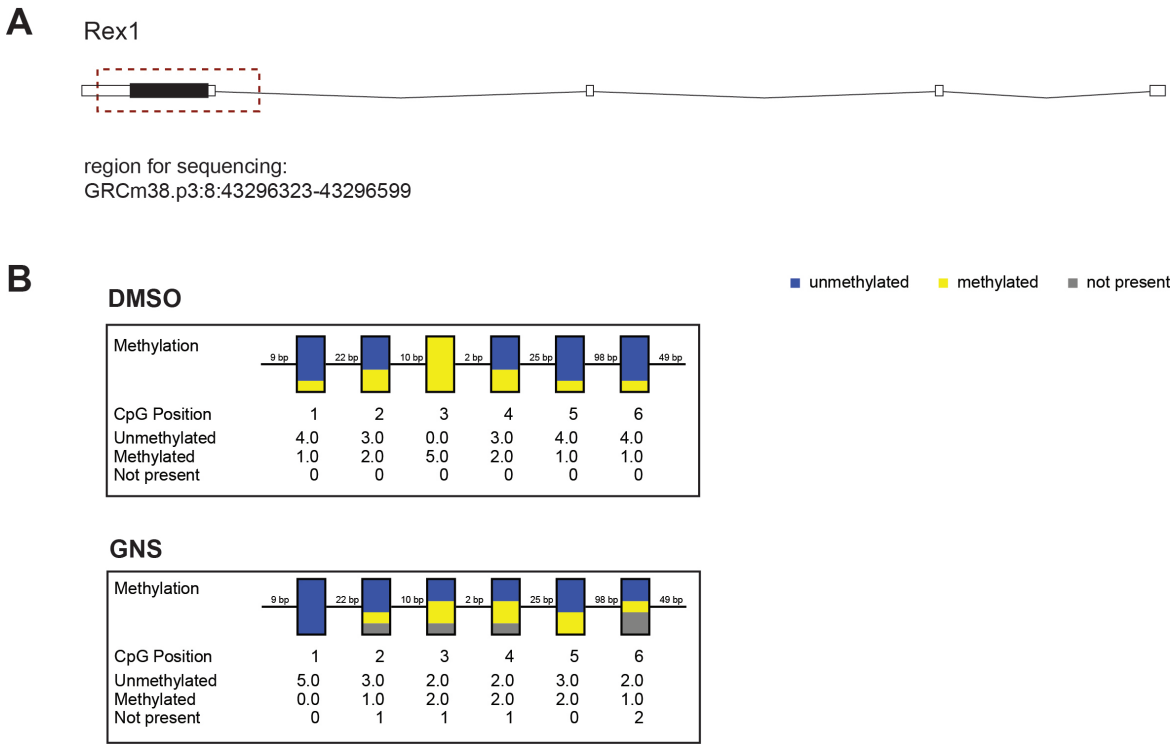

Figure (4.8) Bisulfite sequencing of Rex1. A: Illustration of the region investigated by bisulfite sequencing. B: Methylation of CpG islands within the selected region in DMSO (upper panel) and GNS G (lower panel).

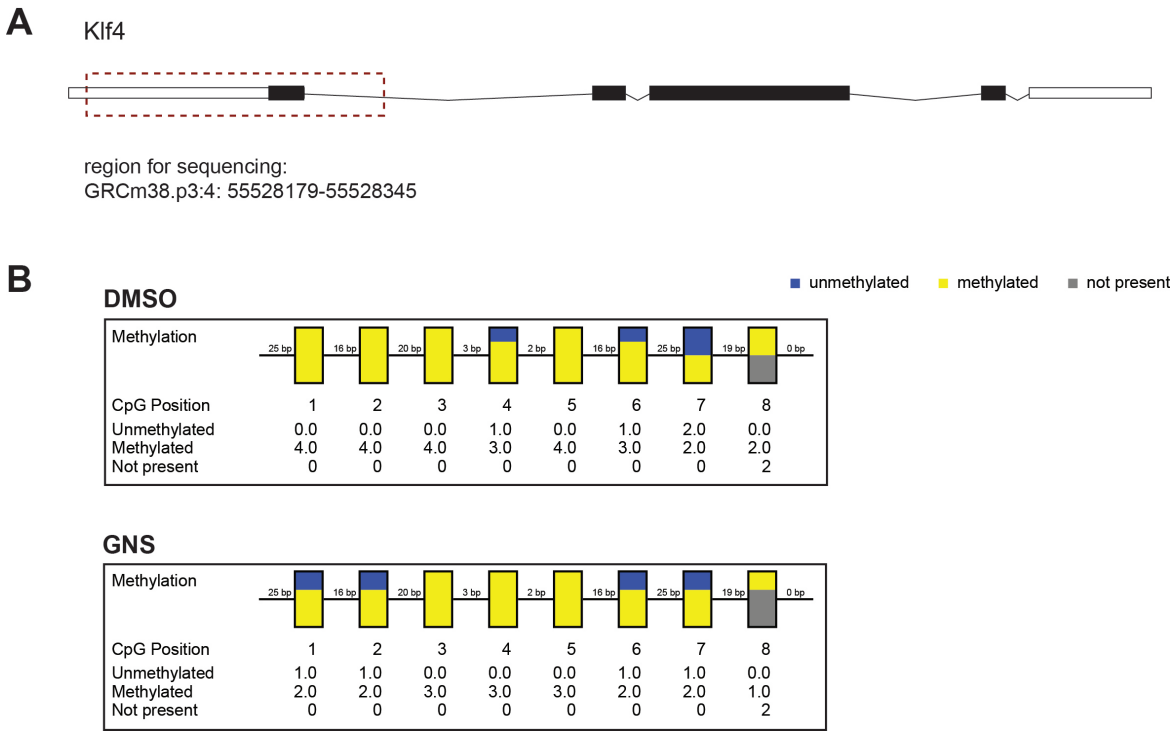

Figure (4.9) Bisulfite sequencing of Klf4. A: Illustration of the region investigated by bisulfite sequencing. B: Methylation of CpG islands within the selected region in DMSO (upper panel) and GNS G (lower panel).

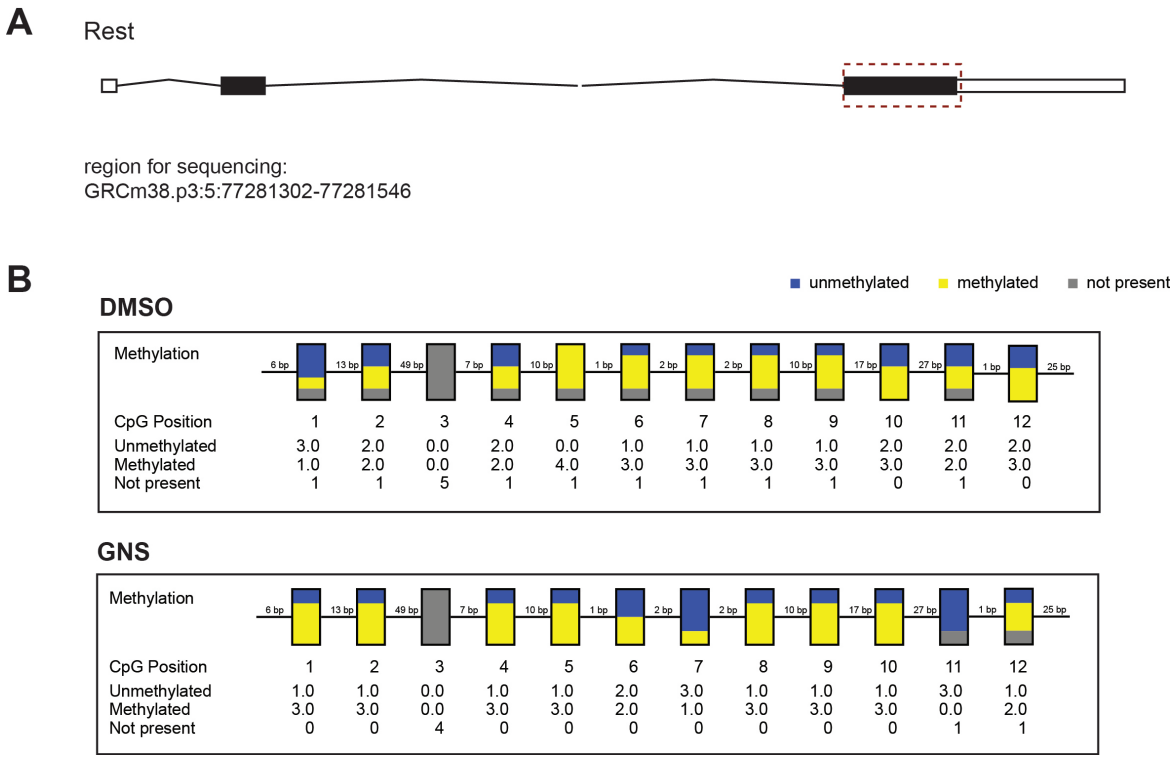

Figure (4.10) Bisulfite sequencing of Rest. A: Illustration of the region investigated by bisulfite sequencing. B: Methylation of CpG islands within the selected region in DMSO (upper panel) and GNS G (lower panel).

control.

The data depicted in figure 4.7 illustrates that the investigated methylation sites on the gene body of *beta-actin* are largely unmethylated which corresponds to active transcription of this gene. No significant changes can be observed in response to GNS G treatment. The pluripotency-related gene *Rex1* (figure 4.8) exhibits a comparable behaviour with overall low methylation in both DMSO and GNS G treated cells. Surprisingly, the second pluripotency gene that was investigated, *Klf4*, appears to be highly methylated at the investigated sites (compare figure 4.9) with no significant changes when cells were treated with GNS G. It remains unclear why Klf4 exhibits such high levels of methylation as the gene is actively transcribed in ES cells. However, it could be speculated that this methylation status is restricted to the respective site

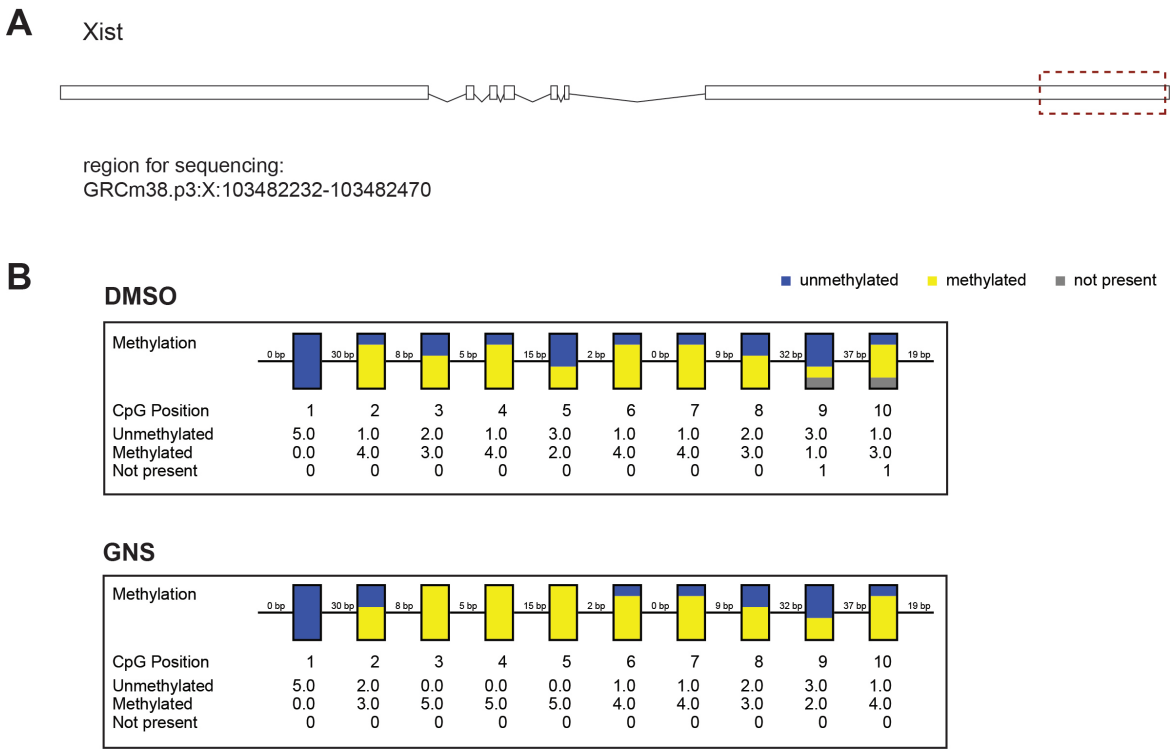

Figure (4.11) Bisulfite sequencing of Xist. A: Illustration of the region investigated by bisulfite sequencing. B: Methylation of CpG islands within the selected region in DMSO (upper panel) and GNS G (lower panel).

and could be overridden by low DNA methylation levels on other exons or promoter regions. When investigating the DNA methylation status for *Rest* which is expressed at very low level in mouse ES cells, it becomes clear that DNA methylation of this gene is also not affected by GNS G treatment (see Figure 4.10). As expected, *Rest* is strongly methylated which corresponds to its silenced state. Investigation of the methylation status of the lncRNA *Xist* indicates no changes in GNS G treated cells in comparison to the control (Figure 4.11). Similar to *Rest*, *Xist* is also heavily methylated in mouse ES cells.

Taken together the results obtained by bisulfite sequencing of selected candidate genes suggest that GNS G does not affect DNA methylation on gene bodies in the respective cases.

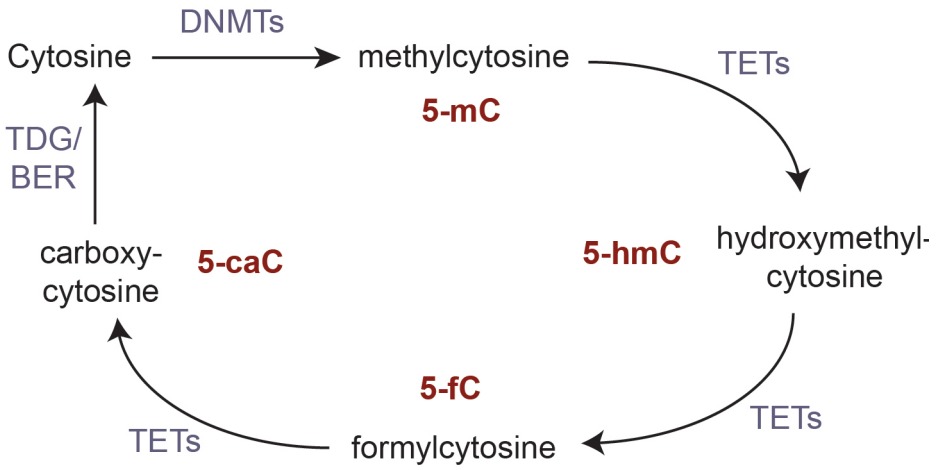

Figure (4.12) Schematic of DNA demethylation. DNA methylation is achieved by addition of a methyl group to cytosine (5-mC). In the progress of demethylation Tet proteins mediated conversion of 5-mC to hydroxymethylcytosine (5-hmC), formylcytosine (5-fC) and carboxycytosine (5-caC). During final step of DNA demethylation 5-caC is reverted to cytosine through thymine DNA glycosylase (TDG)-dependent base excision repair (BER).

4.5.2. Effects of GNS G on global DNA methylation in mouse ES and human iPS cells

The results obtained by bisulfite sequencing indicate no effect of GNS on overall methylation of the gene bodies of the analysed genes. Nevertheless, this analysis does not allow to distinguish between different DNA methylation variants that occur as intermediate during DNA methylation. The variants that have been characterised are 5-methylcytosine (5-mC), 5-hydroxymethylcytosine (5-hmC), 5-formylcytosine (5-fC) and 5-carboxymethylcytosine (5-caC). In order to investigate if GNS G treatment could shift the ratios between those methylation variants, we initially tried to analyse DNA from GNS G treated E14 mouse ES cells and ChiPS4 human iPS cells using dot blots. However, there were technical issues using this method and we found it might not entirely suitable for obtaining reliable results in our hands. With the aim of overcoming those limitations, it was decided to perform a more quantitative measurement

of DNA methylation by HPLC-MS.

The combination of liquid chromatography (HPLC) and mass spectrometry (MS) is a very sensitive tool for the quantification of nucleotides (modified and unmodified). For this experiment, HPLC-MS was used for the quantification of 5-mC and 5-hmC content of DNA from E14 mouse ES cells and ChiPS4 human iPS cells after GNS G treatment.

As can be seen from Figure 4.13 A there is no significant difference in the 5-mC content of E14 cells that had been treated with the inhibitor in comparison to the respective control. Treatment with 2i results in the expected decrease in 5-mC in those cells. Regarding the 5-hmC quantification it can again be observed that GNS G has no effect and levels are comparable to those of the DMSO control (Figure 4.13 B).

The quantification of 5-mC and 5-hmC content in ChiPS4 DNA also shows no differences between cells that had been treated with GNS G and the control (see Figure 4.14). It should be noted that the overall 5-hmC content in ChiPS4 is lower than in E14 cells.

Taken together these results clearly demonstrate that treatment with GNS G does not impact on 5-mC or 5-hmC content of E14 mouse ES cells or ChiPS4 human iPS cells. It is therefore unlikely, that the previously observed effects on gene expression is due to a global change in DNA methylation.

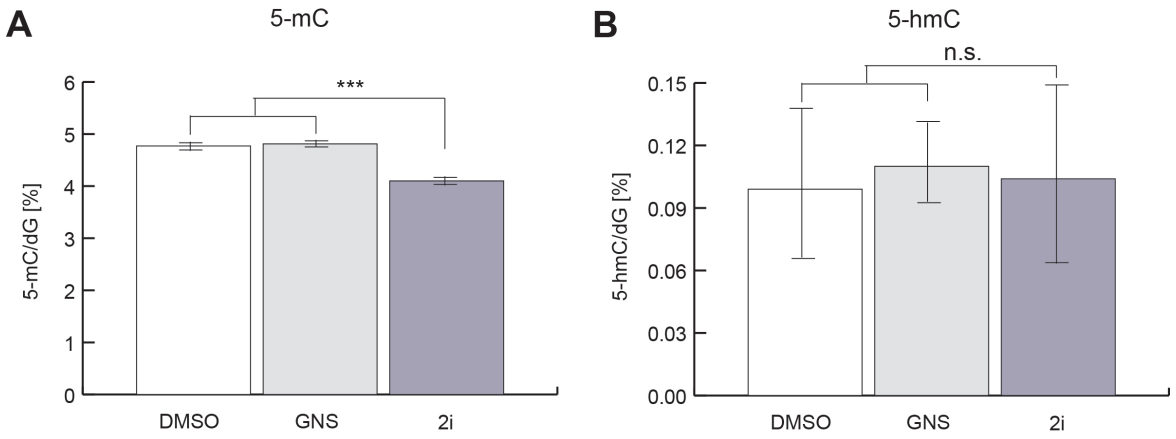

Figure (4.13) Quantification of cytosine methylation of DNA from E14 mouse ES cells by HPLC-MS. A: Quantification of 5-methylcytosine as percentage of total cytosine. B: Quantification of 5-hydroxymethylcytosine. n=8 independent biological repeats for all, ***: $p \leq 0.0001$, n.s.: $p > 0.05$ (Student's t-test, paired), errorbars represent SEM..

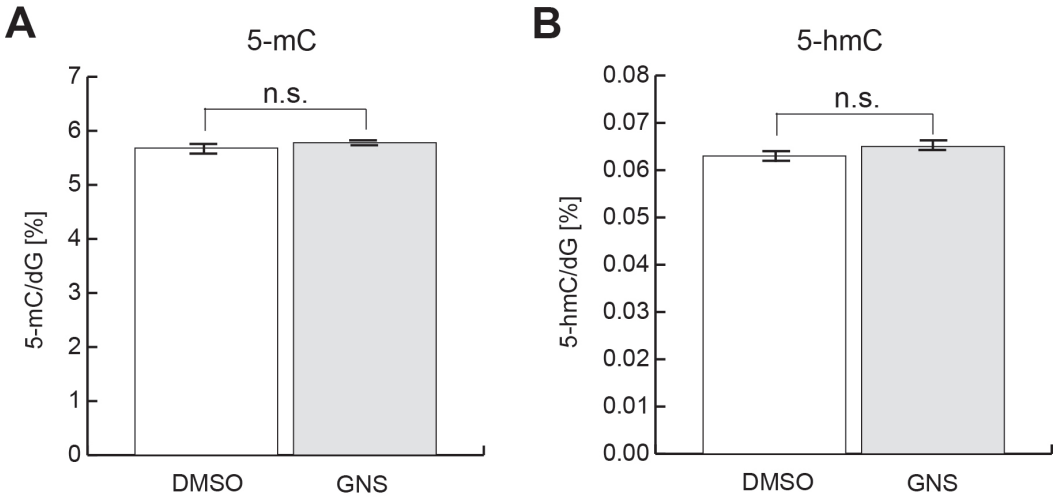

Figure (4.14) Quantification of cytosine methylation of DNA from ChiPS4 human iPS cells by HPLC-MS. A: Quantification of 5-methylcytosine percentage of total cytosine. B: Quantification of 5-hydroxymethylcytosine. n=8 independent biological repeats for all, n.s.: $p > 0.05$ (Student's t-test, paired); errorbars represent SEM.

5. Effects of high O-GlcNAc on gene expression of human iPS cells

The effects of Oga inhibition by GNS G treatment on ectoderm differentiation of human iPS cells described in chapter 3 represent a striking phenotype that has not been described before. Yet the detailed mechanisms that lead to the observed hindrance of ectoderm differentiation remain elusive. Therefore it was decided to perform a messenger RNA (mRNA) sequencing experiment on undifferentiated ChiPS4 human iPS cells with the aim of elucidating the effects of GNS G treatment on gene expression in detail. Results of the sequencing and subsequent bioinformatic analysis will be described in the following sections.

5.1. GNS G induces a feedback regulation of gene expression of Oga and Ogt

In advance to sending samples for RNA sequencing the effect of GNS G was validated by investigating the expression of two known target genes: Ogt and Oga. Previous experiments had demonstrated a feedback regulation of protein levels of the two O-GlcNAc regulating enzymes in GNS G. Hence, it would be expected that a similar effect could be observed on mRNA level. Therefore, ChiPS4 were treated with GNS

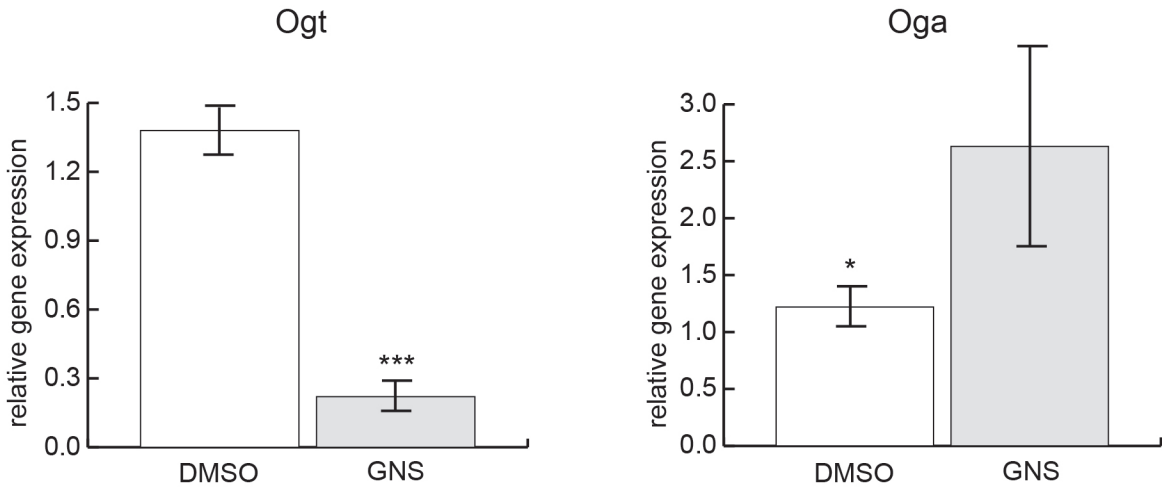

Figure (5.1) Expression of Oga and Ogt after three days of GNS G treatment. Gene expression analysed by RT-qPCR prior to sending samples for RNAseq. n=6 independent biological repeats. ***: $p \leq 0.001$; **: $p \leq 0.01$; *: $p \leq 0.05$; n.s.: $p > 0.05$ (Student's t-test, paired), errorbars represent SEM. (Results for individual experiments are shown in Figure A.1 and Figure A.2.)

G for three days, RNA was extracted, reversely transcribed and analysed for the expression of Ogt and Oga by RT-qPCR. In total, seven independent experiments were performed. Single results for each experiment are given in section A.1. Based on these results, experiment number six was excluded as it the expected feedback was weaker than in the other experiments performed. Combined results of experiments one to five and seven are depicted in Figure 5.1. As can be seen from these data, GNS G treatment induces a pronounced downregulation of Ogt mRNA. On the other hand Oga expression increases. However, it should be noted that the increase in Oga mRNA is a lot more variable between different experiments, whereas the order of magnitude of feedback downregulation of Ogt expression is very consistent between experiments. Overall these results represent an initial validation of the GNS G treatment in regard to its effect on the expression of known targets and RNA samples of experiment one to five and seven were submitted to RNA sequencing.

5.2. Analysis of gene expression in high O-GlcNAc by RNA sequencing

5.2.1. Quality control of the RNA sequencing

In order to determine the quality of the sequencing results analysis of the reads was performed using FastQC (Andrews, 2010). Depicted in Figure 5.2 are representative results of the FastQC analysis focussing on sequencing quality and GC content of the samples. From Figure 5.2 A it can be seen that the quality of the sequencing was very good as throughout the sequencing all reads reached high quality scores (> 28). Importantly, those results were almost identical for GNS G treated samples and the respective DMSO control. Figure 5.2 B illustrates the results of the FastQC analysis for GC content per sequence. The mean GC content should ideally resemble a normal distribution (indicated by the blue line in the graph) and discrepancies can indicate contaminations. In this case, a slight shift of the curve towards a higher GC content can be observed. This indicates a systematic bias which is independent of the base position. As this observation can be made for both GNS G and DMSO treated samples, the curve shift merely indicates a generic feature of the library as the GC content of the experimental library varies from the reference distribution applied by the FastQC module. Overall, the results of the FastQC analysis demonstrate that the sequencing was successful and the data obtained is of high quality.

5.2.2. Differential gene expression

The obtained sequencing data was analysed for genes that are differentially expressed in GNS G when comparing to the DMSO samples. In brief, sequences were aligned to

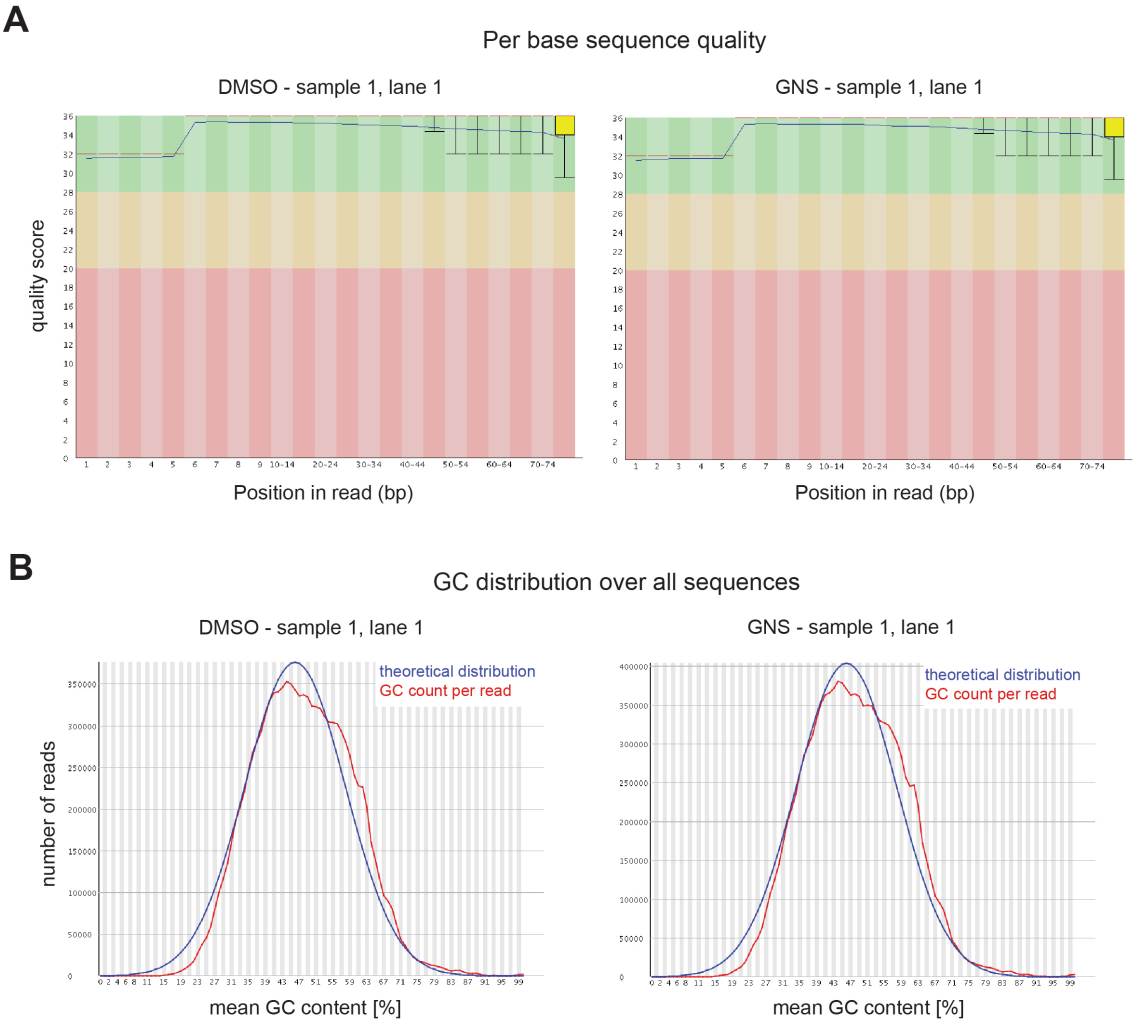

Figure (5.2) FastQC analysis results. **A**: Sequence quality analysis with FastQC. Red central line: median value. Yellow boxes represent the inter-quartile range. Blue line indicates the mean quality. **B**: Per sequence GC content. The blue line indicates a theoretical reference distributions, the experimental distribution is depicted in red.

the human reference genome (GRChg38) using STAR 2.5 (Dobin et al., 2013) and differential gene expression analysis was performed using EdgeR (Robinson et al., 2010). Cut-off values used to generate a list of differentially expressed genes were set according to the values obtained for the known target genes Ogt and Oga. Respectively, genes are considered significantly differentially expressed when the log fold change (logFC) is greater than or equal to 0.25 and the false discovery rate (FDR) is smaller than 0.01. The distribution of differentially expressed genes analysed by EdgeR is illustrated in Figure 5.3. In total, of 60,975 genes were identified. Of those genes 592 genes were significantly upregulated in GNS G treated cells, while 341 genes were downregulated. The volcano plot illustrates that while a high number of genes are differentially expressed, the log fold changes for these genes are low. This indicates that the expression of a large number of genes changes to a small extent in GNS G treated cells. In addition, sample relations were explored by multidimensional scaling using EdgeR's plotMDS. As can be seen from Figure 5.3 B, samples cluster together by biological coefficient of variation (BCV) according to treatments (DMSO or GNS G). This illustrates again that there are significant differences in gene expression between the two treatments.

The full result of the differential gene expression (DGE) is given in table A.1.1. A shortened list with genes with the highest log fold change ($|\text{logFC}|>1.3$) in this analysis is given in table 5.2.2. Positive fold changes indicate genes that have higher expression in GNS G than in DMSO. Negative fold changes indicate genes with lower expression in GNS G than in DMSO. As illustrated in table 5.2.2 the genes that are affected by GNS G treatment include a huge variety of different gene types. The list comprises genes coding for transcription factors, growth factors, transmembrane receptors, histone variants, potential oncogenes and also non-coding RNAs. Interestingly, two known

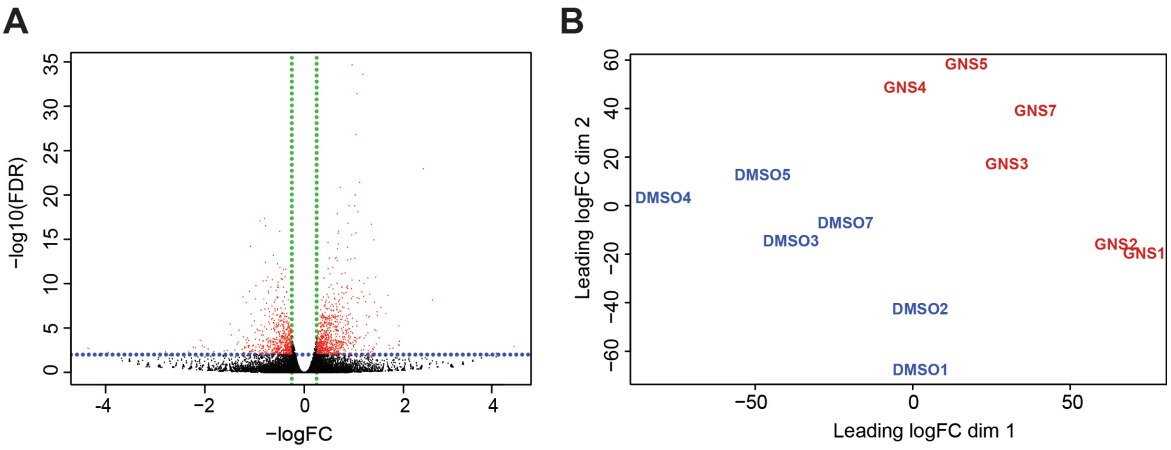

Figure (5.3) Analysis of differential gene expression using EdgeR. A: Volcano plot showing all identified differentially expressed genes. Green lines indicate cut-off for logFC > 0.25. Blue line indicates FDR cut-off (FDR < 0.01), green line indicates cut-off for log fold changes. Genes that were considered significantly differentially expressed are depicted in red. **B**: Multidimensional scaling plot of the fold-changes of biological coefficient of variation (BCV) using EdgeR's plotMDS function.

pluripotency-related genes are found in this list. *Klf2* has been described to sustain a naïve-state pluripotency in human ES cells (Takashima et al., 2014) and is also known to be a Wnt target gene (Qiu et al., 2015). *Fbxo15* is a transcriptional target of Oct4 that is specifically expressed in the early embryo and in mouse ES cells, however, it is dispensable for self-renewal (Tokuzawa et al., 2003). The broad variety of this result indicates a global effect of GNS G and implies that deregulation of O-GlcNAcylation impacts on a multitude of genes involved in many different cellular processes.

Table (5.1) Differentially expressed genes with EdgeR (logFC>1.3)

	external gene name	description	logFC	FDR
1	RP11-261N11.8		-4.22	0.00
2	GALNT9	polypeptide N-acetylgalactosaminyltransferase 9	-2.59	0.00
3	NPTX1	neuronal pentraxin I	-2.40	0.00
4	ESPNP	espin pseudogene	-1.92	0.00
5	KLF2	Kruppel-like factor 2	-1.92	0.00
6	FEZF1	FEZ family zinc finger 1	-1.91	0.00
7	RP11-482M8.1		-1.90	0.00
8	MYADML2	myeloid-associated differentiation marker-like 2	-1.79	0.00
9	LHX5	LIM homeobox 5	-1.72	0.00
10	TMEM88B	transmembrane protein 88B	-1.71	0.00
11	GNA15	guanine nucleotide binding protein (G protein), alpha 15 (Gq class)	-1.69	0.00
12	ZNF781	zinc finger protein 781	-1.65	0.00
13	FEZF1-AS1	FEZF1 antisense RNA 1	-1.55	0.00
14	RAB33A	RAB33A, member RAS oncogene family	-1.51	0.01
15	SPEF1	sperm flagellar 1	-1.51	0.00

16	CALY	calcyon neuron-specific vesicular protein	-1.48	0.00
17	MIAT	myocardial infarction associated transcript (non-protein coding)	-1.47	0.00
18	FBXO15	F-box protein 15	-1.44	0.00
19	ATP6AP1L	ATPase, H+ transporting, lysosomal accessory protein 1-like	-1.44	0.00
20	CABP7	calcium binding protein 7	-1.42	0.00
21	TMEM151A	transmembrane protein 151A	-1.41	0.00
22	SLC16A11	solute carrier family 16 member 11	-1.41	0.00
23	LINC01234	long intergenic non-protein coding RNA 1234	-1.41	0.00
24	CCDC184	coiled-coil domain containing 184	-1.39	0.00
25	TTBK1	tau tubulin kinase 1	-1.37	0.00
26	RAB3C	RAB3C, member RAS oncogene family	-1.36	0.00
27	ASTL	astacin-like metallo-endopeptidase (M12 family)	-1.34	0.00
28	MAPK8IP2	mitogen-activated protein kinase 8 interacting protein 2	-1.32	0.00
29	DRD4	dopamine receptor D4	-1.31	0.00
921	RASSF9	Ras association (RalGDS/AF-6) domain family (N-terminal) member 9	1.31	0.00
922	SLC40A1	solute carrier family 40 (iron-regulated transporter), member 1	1.33	0.00
923	HIST1H2BA	histone cluster 1, H2ba	1.35	0.00
924	FOLR1	folate receptor 1 (adult)	1.39	0.00
925	CITED1	Cbp/p300-interacting transactivator, with Glu/Asp rich carboxy-terminal domain, 1 [Source:HGNC Symbol;Acc:HGNC:1986]	1.47	0.00
926	AC007682.1		1.49	0.00
927	VIT	vitrin	1.61	0.00
928	KDM4E	lysine (K)-specific demethylase 4E	1.93	0.00
929	GDF6	growth differentiation factor 6	2.05	0.00
930	TNFSF10	tumor necrosis factor superfamily member 10	2.09	0.00
931	HIST1H2AA	histone cluster 1, H2aa	2.09	0.00
932	TFEC	transcription factor EC	2.33	0.01
933	PHKA1-AS1	PHKA1 antisense RNA 1	4.35	0.00

5.2.3. Gene ontology analysis

With the aim of determining underlying processes and pathways that are influenced by GNS G treatment the previously obtained list of differentially expressed genes was used as an input for further analysis. Gene ontology (GO) enrichment analysis provides a tool for assessing in which biological processes the respective genes and their products are involved. Gene sets are associated with GO terms that underlie a hierarchical system. The three root terms are "cellular component", "biological process" and "molecular function". Each of those root terms holds a large number of subordinated terms which increase in specificity and therefore give hints to the affected cellular mechanisms.

Overall, GO analysis indicates that the effects of GNS G treatment are widely spread and a large number of GO hits are found in all three root terms. Selected results of the GO analysis for the two root terms "biological process" and "cellular component" are described in the following. Full results of GO analysis are given in the appendix (table A.1.2, table A.1.2, table A.1.2).

The analysis on biological processes revealed that some important feedback mechanisms of O-GlcNAcylation are affected by GNS G treatment (see table 5.2). The GO term hits include "negative regulation of phosphorylation" and "negative regulation of protein phosphorylation". These hits are most likely linked to the competition between O-GlcNAcylation and phosphorylation and therefore represent an important control of the drug treatment with the Oga inhibitor. In addition, three GO terms come up in this analysis that are linked to nutrition and the response to nutrient levels. As O-GlcNAc functions as a metabolic sensor this is another major mechanism that would have been expected to be affected. Therefore, these results provides an

important control and underline the reliability of the data obtained.

Table (5.2) GO Process

GO term ID	Description	Proteins	Hits	p-value
GO:0042326	negative regulation of phosphorylation	359	26	0.00092
GO:0001933	negative regulation of protein phosphorylation	326	24	0.00114
GO:0031669	cellular response to nutrient levels	183	16	0.00131
GO:0031667	response to nutrient level	397	30	0.00019
GO:0007584	response to nutrient	164	16	0.00040

When investigating in which cellular compartments the products of genes affected by GNS G treatment are localised, it becomes clear that the effect is widely spread throughout the cell (compare table A.1.2). Enrichment analysis for the GO parent term "biological process" revealed that products of genes that are differentially expressed in GNS G are also involved an a large number of varied biological processes (compare table A.1.2). Figure 5.4 illustrates these findings and provides an overview of the multitude of GO processes that are affected by GNS G treatment according to GO enrichment. This graph was obtained using REViGO for reduction and visualization of GO results (Supek et al., 2011). In the light of the previously described data on ectoderm differentiation it is of particular interest that many GO term hits are closely related to central nervous system and brain development as well as to neurogenesis. These findings indicate that O-GlcNAcylation might not only play a role in ectodermal differentiation, but it could potentially impact strongly on early neural development. In addition, two hits were identified for GO terms related to growth factor response. The appropriate response to growth factor stimuli is essential for driving differentiation and embryonic development. It might therefore be possible that O-GlcNAcylation hinders differentiation by impairing the cell's reaction to relevant factors.

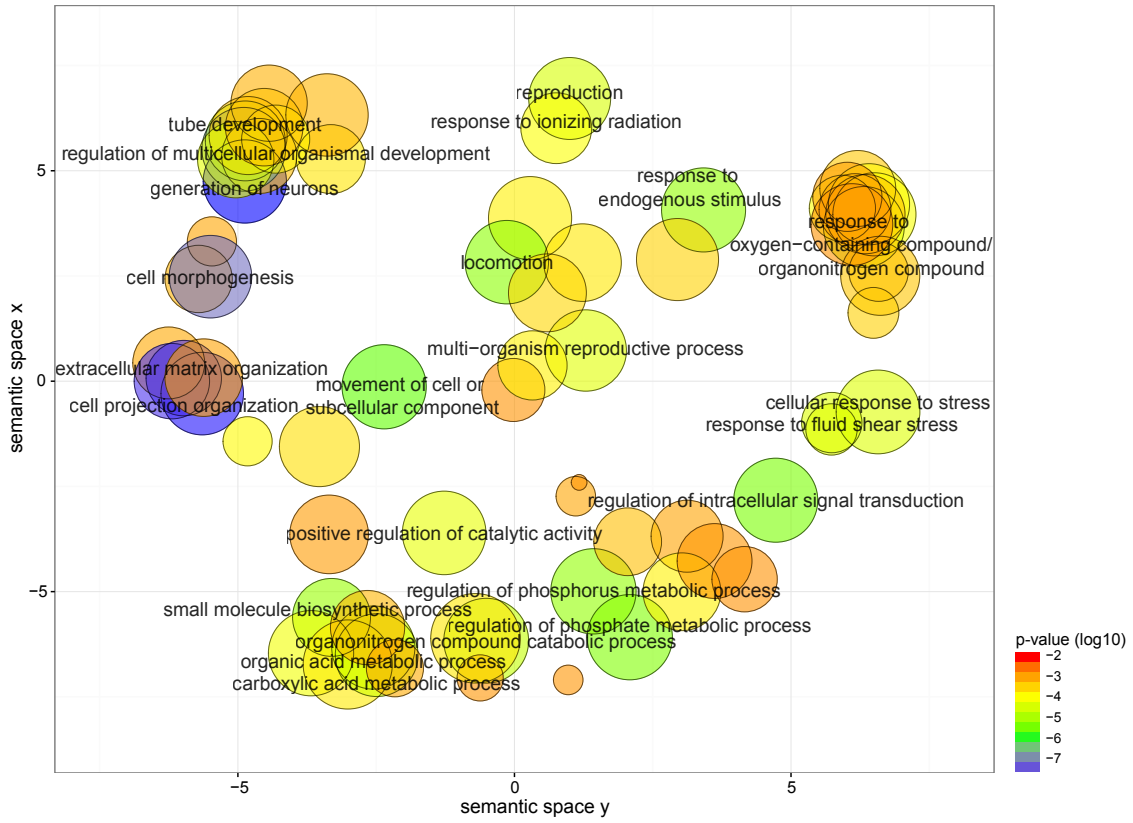

Figure (5.4) Visualization of gene ontology results using REViGO. REViGO analysis indicates a broad effect of GNS G treatment on GO process terms. Colours correspond to log10 p-values of the previously performed GO analysis, circle sizes are correlated to the frequency of the GO term in the underlying GOA database (more general terms are shown as larger circles).

5.2.4. Interaction network analysis

Another tool for extending the information of differential gene expression analysis is StringDB (Franceschini et al., 2013). It allows investigations into interactions of products of the genes that are differentially expressed. A network of predicted protein interactions is build based different levels of evidence, i.e. textmining, database or textmining evidence. For this analysis, the differentially expressed gene list was restricted to the genes that had an absolute log fold change of 0.85 or higher.

The results of StringDB is shown in Figure 5.5. The networks illustrates that a large number of proteins that are affected by GNS G treatment could potentially interact. Overall, 307 possible interactions were detected between the 200 proteins that were analysed. It becomes clear that there are a large proportion of the depicted interactions are based on neighbourhood evidence (green lines) only and the respective proteins very often only show a single line of connection.

There appear to be two clusters of interacting proteins that are both highly deregulated by GNS G treatment and also strongly connected by various lines of evidence. The smaller cluster contains (amongst others) the proteins BMPR1B, GDF6, TNFSF10, CITED1, LEFTY1 and GDNF. Within the cluster especially BMPR1B is connected by experimental evidence to LEFTY1 and GDF8. This smaller cluster is again linked to a second, bigger cluster which contains for example KLF2, CDK5, ZIC5, FOXJ1, MAPK11 and FEZF1. Within this cluster, proteins are connected by different lines of evidence. On the periphery of the second cluster, two histone variants (HIST1H2BA and HIST1H2AA), two zinc finger proteins are found (ZNF781, ZNF436) and two transcription factors (TFEB, TFEC) are found. This might possibly suggest an effect on transcriptional regulation. However, it is difficult to draw

conclusions with a view to the molecular function of high O-GlcNAc levels from this data. The interaction network demonstrates that the effects of GNS G affect a large number of proteins with different functions and there is a high likelihood for crosstalk between them.

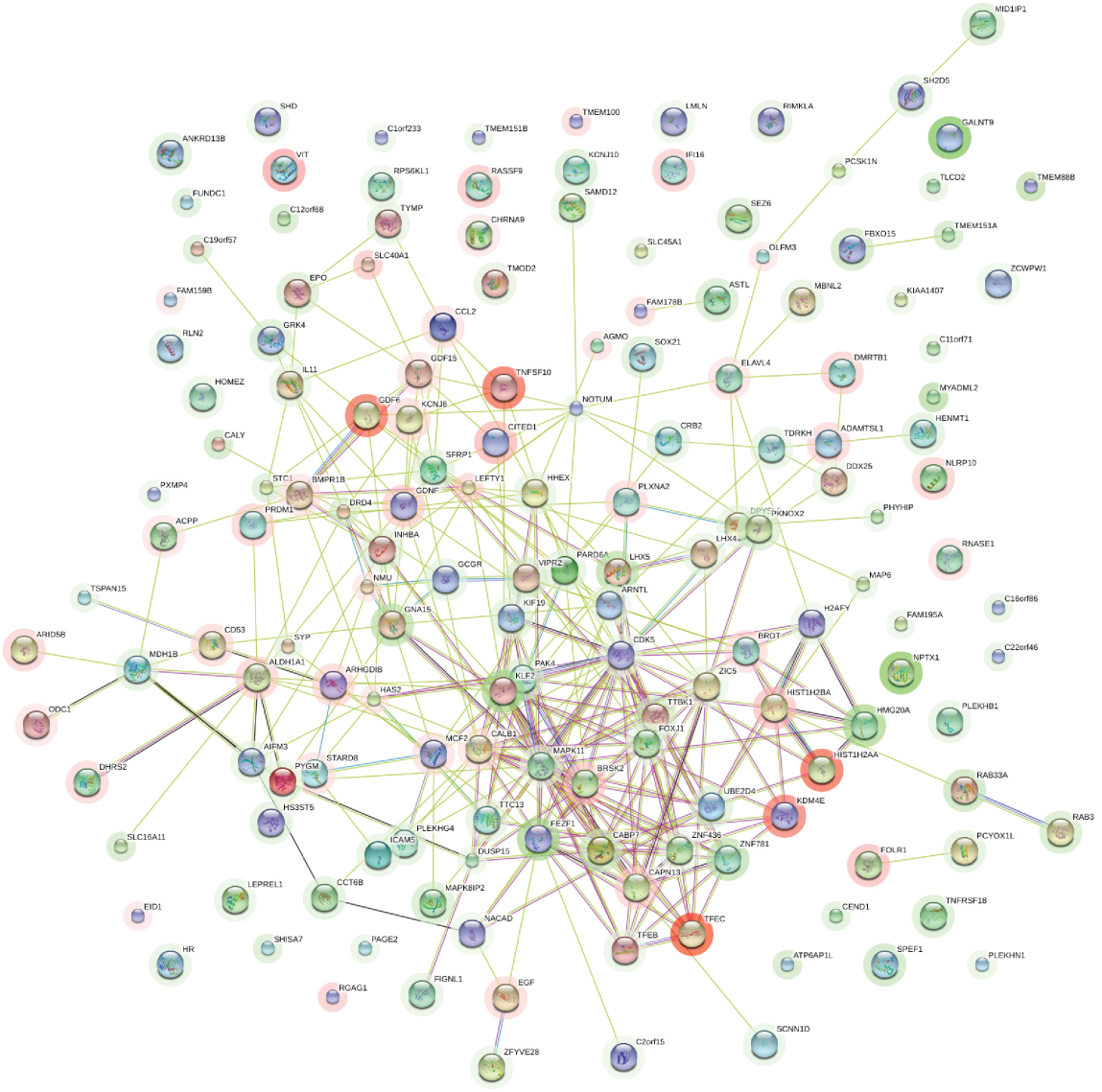

Figure (5.5) StringDB protein interaction network. Spheres represent the respective proteins, halos around spheres indicate the degree of up- (green) or downregulation (red) in GNS G. Lines represent evidence of connection: Red line - fusion evidence; Green line - neighbourhood evidence; Blue line - co-occurrence evidence; Purple line - experimental evidence; Yellow line - textmining evidence; Light blue line - database evidence; Black line - co-expression evidence.

5.3. Validation of selected hits of DGE analysis

In order to validate the data obtained by RNA sequencing with another method, RT-qPCR was performed. Therefore, a set of genes were selected that would span a range of different log fold changes with the aim of analysing the correlation between both methods. As representative genes with a high log fold change *Klf2* and *GDNF* were analysed. *FoxJ1* and *Sema3A* represent genes with a intermediate log fold change whereas *MAPK21* and *Klf8* have a small log fold change.

The respective results are illustrated in Figure 5.6. The data obtained by RNA sequencing is in most cases closely mimicked by the RT-qPCR data. Genes that were detected with a high log fold change in the differential gene expression analysis also show a significant difference by RT-qPCR (i.e. *Klf2*). Interestingly, for *GDNF* and *Sema3A* the results from RNA sequencing do not quiet correlate with those obtained by RT-qPCR: While *GDNF* has a higher log fold change than *Sema3A*, *Sema3A* appears to be more upregulated by RT-qPCR. It is also important to note that the two genes with small log fold changes (around $|0.35|$), namely *MAPK21* and *Klf8*, still show significant up- or downregulation by RT-qPCR. Altogether, these data support the results of the RNA sequencing experiment very well and indicate that in most cases RT-qPCR data and differential gene expression data correspond closely.

5.4. GNS G treatment interferes with signal transduction of the BMP signalling pathway

From the results obtained by differential gene expression analysis several hits stood out that were linked to the BMP signalling cascade. An overview of those hits and their

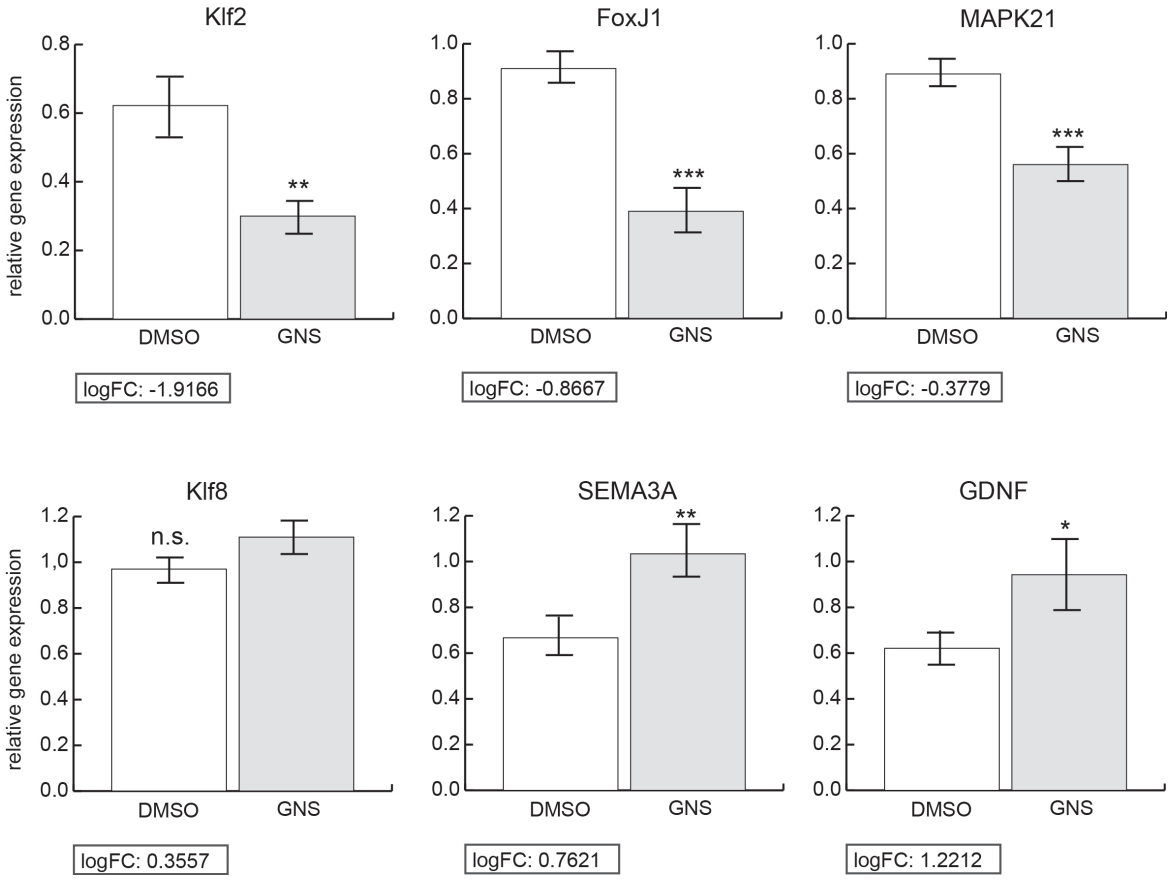

Figure (5.6) Validation of differential gene expression results. RT-qPCR analysis of gene expression on selected hits obtained by differential gene expression analysis. ***: $p \leq 0.001$; **: $p \leq 0.01$; *: $p \leq 0.05$; n.s.: $p > 0.05$ (Student's t-test, paired); n=6 independent biological repeats for all, errorbars represent SEM.

function within the BMP pathway is depicted in Figure 5.7 C (hits are highlighted in green). First, the gene coding BMP receptor subtype BMPR1B (also known as Alk6) is upregulated in GNS G treated ChiPS4 cells. In addition, the inhibitory *Smad7* is upregulated in GNS G in the differential gene expression analysis. Furthermore, the gene *TGFBI* (TGF-β induced) exhibits an increase in expression in GNS G treated cells. *TGFBI* is linked to the BMP signalling pathway via the crosstalk between Smads and TGF-β signalling.

In order to investigate whether BMP4 signalling was modified in GNS G treated human iPS cells, the findings of the RNA sequencing experiment were first validated by RT-qPCR (Figure 5.7 A). Both *BMPRB1* and *TGFBI* show a strong increase in gene expression in response to GNS G treatment which validates the previous findings. For *Smad7*, a trend towards upregulation of gene expression in GNS G could be observed. However, this result is not statistically significant.

The increase in expression of the BMP receptor subtype BMPRB1 might imply that the cells are more susceptible to BMP4 stimulation. In the interest of testing this hypothesis, the pathway was stimulated with BMP4 and the expression level of *Cdx2*, a transcriptional target of the BMP4 cascade, were determined by RT-qPCR. ChiPS4 were treated with GNS G for three days prior to overnight stimulation with different concentrations of BMP4. The respective result from RT-qPCR is shown in Figure 5.7 B. The data demonstrate a pronounced increase in expression levels of *Cdx2* in GNS G treated cells in comparison to the untreated and DMSO control. This effect can already be observed at very low concentrations of BMP4 and even without BMP4 stimulation. The result suggests that the BMP4 signalling pathway might become more active in GNS G treated cells. As the BMP4 pathway has an inhibitory impact on neural differentiation and embryogenesis in various model systems (LaVaute

et al., 2009; Khokha et al., 2005; Wilson et al., 1997; Dosch et al., 1997) it might be speculated that its increased activity could hinder differentiation and therefore provide a molecular mechanism for the previously described effects on neural differentiation. However, it needs to be noted that this result is preliminary and future work should therefore be targeted on validating this data and investigating further effects on BMP4 signalling.

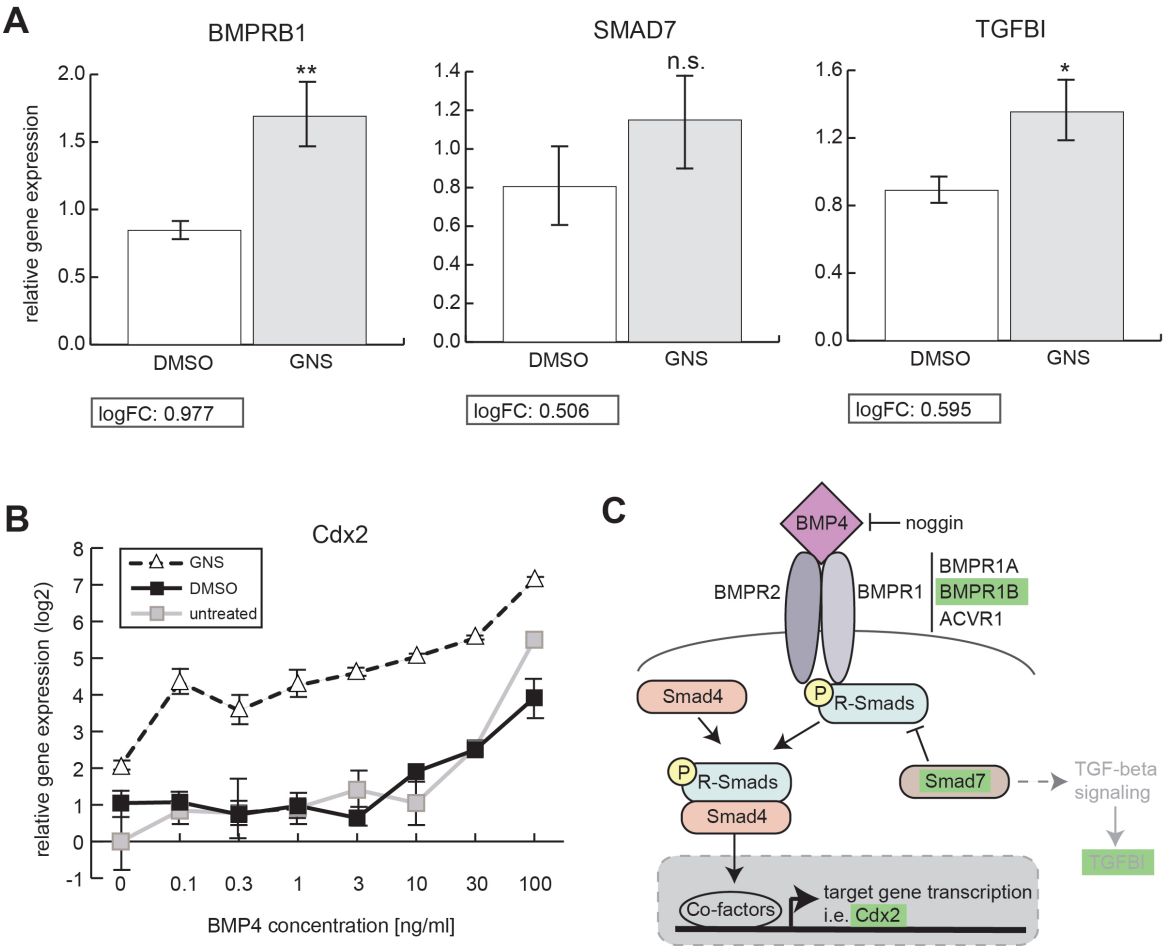

Figure (5.7) GNS G treatment affects the BMP4 signalling pathway in ChiPS4. A: RT-qPCR validation of differential gene expression hits related to BMP4 signalling; **: $p \leq 0.01$; *: $p \leq 0.05$; n.s.: $p > 0.05$ (Student's t-test, paired); n=6 independent biological repeats for all, errorbars represent SEM. **B**: RT-qPCR analysis of the BMP4 target gene *Cdx2* in response to stimulation with BMP4; n=1 biological repeat, errorbars represent technical error of the RT-qPCR (3 technical repeats of RT-qPCR). C: Schematic of the BMP4 signalling cascade, hits from differential gene expression analysis are highlighted in green.

6. Discussion

6.1. Results summary

The aim of this study was to investigate the role of O-GlcNAc on pluripotent stem cell differentiation into different lineages and to explore the molecular mechanisms of any effects discovered. These findings would then provide valuable starting points for future research in the field and contribute to the overall understanding of the molecular functions of this posttranslational modification in pluripotent stem cells, lineage commitment and early mammalian development.

The findings presented in the previous chapters illustrate that O-GlcNAc plays a vital role in the differentiation of both mouse ES cells and human iPS cells. Elevated O-GlcNAc levels impair differentiation in both pluripotent stem cell models. Importantly, this effect is lineage independent in mouse ES cells, but restricted to ectodermal differentiation in human iPS cells. Experiments on human iPS cells have shown that the early stages of lineage commitment are strongly affected by deregulation of O-GlcNAcylation. At advanced stages, elevated O-GlcNAc levels do not impair ectodermal differentiation, neurogenesis or the maintenance of neurons.

With a view on potential molecular mechanisms underlying the described effect on differentiation, several approaches were taken. For mouse ES cells, data previously

obtained by microarray analysis suggested that gene sets that are normally silenced get re-expressed upon GNS C treatment. Following up on these data, it was possible to validate the microarray findings on the gene sets of polycomb repressive complex repressed genes and 2C genes, a gene set characteristically expressed during the two-cell-stage of mouse embryonic development. When investigating potential molecular mechanisms that might cause the re-expression of those genes, the focus lay on global mechanisms that could be linked to both O-GlcNAc and transcriptional activation from previous studies. It was found that both histone acetylation and DNA methylation are not affected by high O-GlcNAc levels to a degree that would explain the effect on silenced genes. For human iPS cells, a RNA sequencing experiment was performed to elucidate potential mechanisms that cause the observed hindrance of ectodermal differentiation in high O-GlcNAc conditions. This experiment and the subsequent data analysis revealed that elevated O-GlcNAc levels impact on a multitude of genes. Overall, it could be observed that many genes were affected to a small extent by high O-GlcNAc levels, suggesting a multitude of small effects that orchestrate a global cellular response. The data suggest that underlying mechanisms are complex and most likely interacting on multiple levels. Regardless of those findings, it was possible to reveal one signalling cascade that responds to elevated O-GlcNAc levels by an increase in activity: the BMP pathway. One previous study has described potential crosstalk between O-GlcNAc and BMP signalling, however this has not been linked to pluripotent stem cells and their differentiation before (Herhaus, 2014). These findings provide a good starting point for future studies that could, for example, focus on the link between the function of O-GlcNAc in BMP signalling and neurogenesis in the mammalian embryo.

6.2. GNS G treatment: differences between high O-GlcNAc and Oga inhibition

When evaluating the results represented in this thesis it is essential to keep in mind the nature of Oga inhibitor treatment used for the majority of experiments. Both in mouse ES and human iPS cells treatment with GNS G has multiple effects. First, O-GlcNAc levels get elevated globally as a consequence of inhibition of the hydrolase Oga. In addition, protein levels of the two O-GlcNAc regulators Oga and Ogt are affected by feedback mechanisms. The excess of O-GlcNAcylation results in an upregulation of Oga while its counterpart, the transferase Ogt, is downregulated. Therefore, treatment with GNS G has three distinct effects and each of them could trigger various molecular mechanisms. An increase in O-GlcNAc modification on proteins will most likely impact on their function in cellular signalling. This could happen through multiple mechanisms. For example, O-GlcNAcylated proteins could be functionally distinct from their unmodified counterparts as has been proposed for RNA polymerase II (Comer et al., 2001). Increased O-GlcNAcylation does also impact on phosphorylation as both modifications compete. Although data from our lab suggests that the global elevation of O-GlcNAc levels does not lead to an overall decrease in phosphorylation, it is likely that selected proteins can be affected in such a manner (Bauer et al., 2015; Speakman et al., 2014). The feedback regulation of Oga and Ogt will have further effects on cell signalling that might act in concert or independently from globally increased O-GlcNAcylation. Especially Ogt has been proposed to complex with numerous other enzymes and in many cases the detailed functions of Ogt therein have not been revealed. However, a downregulation of Ogt in response to GNS G might well lead to a decrease in the formation of Ogt-containing regulatory complexes. As an

example, disruption of the previously proposed Ogt/Tet1/2/Sin3A/HDAC1 complex (Vella et al., 2013) could result in a loss of repression on certain genetic elements. For these reasons, it is important that the effects of GNS G treatment are not limited to an increase in O-GlcNAcylation levels. In order to decipher the differences between feedback effects and high O-GlcNAc it would be vital to use a combination of inhibitor treatment, knockdown and overexpression experiments of Ogt and Oga. Unfortunately, there are technical limitations that prevented me from conducting some of those experiments in mouse ES and human iPS cells. In mouse ES cells, transient knockdown of Ogt and Oga could be successfully performed for a short period using siRNAs. However, siRNA-mediated knockdown of Ogt strongly disrupted the viability of mouse ES cells in our hands. Transient overexpression of Oga and Ogt in mouse ES cells was also possible, although the overexpression of Ogt remained at a very low level and might therefore not have a biological effect. Efforts were made to generate inducible knockdown cells from both mouse ES and human iPS cells. In both cases, an efficient knockdown of either Oga or Ogt have not be achieved to date. Consequently, experiments had to be limited to using GNS G treatment in human iPS cells. Where mouse ES cells were used, transient knockdown and overexpression of Oga were conducted during short-term experiments.

6.3. O-GlcNAc in mouse ES cells and human iPS cells: effects on differentiation

With a view to extending current knowledge regarding the role of O-GlcNAcylation in pluripotent stem cells, the data in this thesis has made a significant contribution for both mouse ES and human iPS cells. For mouse ES cells, only an effect of treat-

ment with the hexosaminidase inhibitor PUGNAc on cardiomyocyte differentiation had previously been demonstrated by Kim et al. (2009). Data presented in the study at hand has contributed to extending these observations by showing that all lineages are affected by treatment with GNS G. Further data from our lab demonstrates a prominent effect of elevated O-GlcNAc levels on neural lineage commitment and adds to the significance of the data presented in this thesis (Speakman et al., 2014). However, it remains to be clarified at what stage differentiation and lineage commitment can be disrupted by modified O-GlcNAcylation. The data at hand implies that the transition from a naïve to formative pluripotency is not impaired by GNS G treatment. For example, correct O-GlcNAcylation might only be important for the regulation of lineage-specific transcription factors which are commonly expressed at an advanced stage of differentiation. Consequently, the transition to a formative stage would not be affected, whereas further differentiation would be blocked, possibly through a mechanism controlling transcription of relevant genes and factors. It would therefore be important to elucidate during which stages of differentiation deregulation of O-GlcNAc hinders the correct progression of the cells towards a certain lineage and whether differences between the three lineages could be observed in this aspect. Investigations into the expression profile of naïve, primed and relevant lineage markers would be of particular importance to provide a good understanding on this matter.

Intriguingly, while a lineage-independent effect was observed in mouse ES cells, treatment with the Oga inhibitor GNS G only disrupts differentiation into ectoderm when using human iPS cells. Prior to the experiments conducted for this study, only one report had linked O-GlcNAcylation and differentiation in human pluripotent stem cells: Maury et al. (2013) describe that the maintenance of pluripotent stem cells is not disrupted by treatment with PUGNAc/Thiamet G, an observation that is in agreement

with the data presented in this thesis using GNS G. Using a spontaneous differentiation assay, Maury et al. (2013) concluded that some lineage markers were differently expressed when cells had been treated with PUGNAc/Thiamet G. In particular, they describe downregulation of the ectoderm markers Pax6 and Msx1 and upregulation of two endoderm markers. However, in particular with a view to their results regarding the endoderm markers it should be noted that the apparently significant differences can only be observed at one intermediate timepoint of the differentiation assay. At later stages, PUGNAc/Thiamet G treated cells do not show an impact of inhibitor treatment on the expression pattern of those markers. Although the data presented by Maury et al. (2013) also hints towards an effect of Oga inhibition on ectoderm differentiation the detailed results are not consistent with data presented in this thesis. While Maury et al. (2013) report a prominent effect on *Pax6* gene expression, *Pax6* is the only investigated ectoderm marker that was not deregulated during the experiments presented in the study at hand. It is unclear why *Pax6* appears to be unaffected in GNS G treated cell during ectoderm differentiation while all other neuroectodermal lineage markers that were investigated, i.e. *Sox1*, are strongly downregulated in GNS G treated cells. Data provided by Chambers et al. (2009) suggests that *Pax6* is induced at a slightly later timepoint in differentiation than *Sox1* which could explain the differences in behaviour of those two markers as the effect of GNS G appears to be affecting early timepoints predominantly. However, data presented in this thesis does not support this hypothesis as both markers appear to be expressed in a similar temporal pattern in control cells. Differences in the behaviour of *Pax6* and *Sox1* have also been reported by other groups (Neely et al., 2012), although the reasons have not been elucidated. As an effect of O-GlcNAc has been established in relationship with transcriptional and epigenetic regulation, it might also be speculated that GNS

G treatment leads to repression of selected transcription factors and thereby hinders neuroectodermal lineage commitment. This could explain why some markers are entirely repressed, while others like *Pax6* are induced normally. Furthermore, expression of *Pax6* and *Pax7* is responsive to BMP signalling during development of the neural tube. While *Pax7* is expressed in the dorsal region (high BMP), *Pax6* is found more ventrally (low BMP) (Timmer et al., 2002; Briscoe and Small, 2015). Therefore, increased sensitivity towards BMP4 in GNS might affect *Pax7* expression, but not the induction of *Pax6*.

The discrepancies between the report from Maury et al. (2013) and the data in this thesis could have several potential reasons. Firstly, the data presented by Maury et al. (2013) is limited to spontaneous differentiation assays in EBs. While carrying out similar experiments I found that marker gene expression at final stages of spontaneous differentiation is difficult to investigate in human iPS cells. Due to the characteristics of this type of differentiation the cell population harvested at the endpoint of the assay consists of numerous different cell types and lineages. Analysis of gene expression requires investigating the whole cell population with the result that specific lineage markers are only expressed in a small subset of cells and therefore difficult to detect in meaningful quantities. Secondly, the effects of PUGNAc and Thiamet G might be different from those observed when using GNS G as they are known to also target other hexosaminidases than Oga. Moreover, differences in the final analysis step by RT-qPCR, i.e. primers used, also might contribute to the observed divergence. The investigations presented in this thesis also extend the previous knowledge by providing morphological evidence that demonstrates a strong disruption of *in vitro* neurogenesis. As rosette formation bears close similarity to the establishment of the neural tube, this observation puts further impact on the data and hints that O-GlcNAcylation might

directly affect early developmental steps in the mammalian embryo. Furthermore, it could be clearly demonstrated, that GNS G treatment only disrupts ectoderm differentiation when added at early stages of the process. O-GlcNAcylation might therefore be essential for early lineage decisions, but not necessarily for neuron formation or maintenance once the cells have committed to their ectodermal fate.

6.3.1. Differences in model systems: Mouse ES cells vs. human iPS cells

From the data previously presented it becomes clear that deregulation of O-GlcNAc levels by GNS G treatment has very different effects in mouse ES cells and human iPS cells. Most importantly, GNS G treatment impairs differentiation into all lineages in mouse ES cells but only has an effect on neuroectoderm differentiation of human iPS cells. In addition, differentiation appears to be retarded in mouse ES cells whereas the data at hand suggests that ectoderm differentiation of human iPS cells can be entirely blocked by GNS G treatment. It should also be noted at this point that human iPS cells treated with GNS G loose expression of pluripotency markers (*Oct4, Nanog*) at the same rate as control treated cells. Therefore, the cells appear to become arrested after switching off expression of pluripotency genes, but before expressing lineage markers. The identity of those cells is therefore unclear. In comparison, mouse ES cells treated with GNS G retain *Oct4* expression longer than cells treated with a vehicle control which implies that the complete differentiation process is retarded (Speakman et al., 2014). From these differences in effects of GNS G treatment on differentiation of mouse ES and human iPS cells the question arises whether the discrepancies between those two cell models are species-dependent (mouse versus human) or due to differences in

the pluripotency model (ES versus iPS).

Current knowledge strongly suggests that mouse and human ES/iPS cells represent different developmental stages: While mouse ES cells are in a state of naïve pluripotency, human ES cells bear more similarity with a primed pluripotent state in mouse and can therefore be compared to mouse epiblast stem cells (EpiSCs) derived from the postimplantation embryo (Tesar et al., 2007; Brons et al., 2007). Selected molecular differences between mouse ES and human ES/iPS cells are summarized in table 6.1. Very recently, several reports have claimed the that cells of a naïve pluripotent state can be derived from human embryos (Guo et al., 2016; Warrier et al., 2016). In addition, multiple studies have undertaken approaches to "revert" human ES cells to a state resembling the naïve ground state of mouse ES cells using combinations of multiple cytokines and small molecule inhibitors (Theunissen et al., 2014; Ware et al., 2014; Takashima et al., 2014; Chan et al., 2013; Gafni et al., 2013). Although these studies indicate that human ES/iPS cells can exist in a state that shares most understood molecular and epigenetic characteristics of the murine ground state pluripotency it can be argued to what extent these naïve human ES/iPS cells actually mirror the *in vivo* situation within the human embryo. For example, Pastor et al. (2016) point out that maintaining human ES cells in a naïve state results in the loss of DNA methylation at key imprints and those cells therefore lack the epigenetic memory of the human embryo. An earlier study by Huang et al. (2014) also demonstrates that the transcriptome of naïve human ES cells holds little resemblance with the ground-state mouse ES cell transcriptome. However, both have been shown to clearly correspond to their respective preimplantation epiblasts in this study.

Apart from species-dependent discrepancies, there are also reports on differences between ES and iPS cells (compare section 1.1.3). iPS cells are cultured and differen-

Table (6.1) Differences between mouse ES cells and human ES/iPS cells (Gafni et al., 2013; Marks et al., 2012; Hanna et al., 2010)

Feature	mouse ES cells	human ES/iPS cells
Oct4 expression through	distal enhancer	proximal enhancer
X-chromosome	active	tendency to inactivation
DNA methylation	very low	increased
H3K27me3 on lineage-regulatory genes	not deposited	deposited

tiated under the same conditions as ES cells and have been demonstrated to demonstrate a comparable level of cell potency (Takahashi and Yamanaka, 2006; Takahashi et al., 2007). Regardless, it appears highly plausible that iPS cells can to some degree retain a "somatic memory" after reprogramming which is most likely reflected by differences in DNA methylation patterns between ES and iPS cells (Doi et al., 2009; Kim et al., 2009; Ghosh et al., 2010; Lister et al., 2011).

Both interspecies differences and divergence between ES and iPS cells are found on the level of gene regulation and expression. Numerous of these regulatory processes have reported interaction with O-GlcNAc and its regulators which might be the reason why deregulation of the modification impacts differently on mouse ES and human iPS cells. For example, human ES/iPS cells already show deposition of the histone mark H3K27me3 on lineage-dependent genes which is mediated by PRC2. For recruitment of PRC2, Tet1-mediated demethylation is essential (Wu and Zhang, 2011; Aloia et al., 2013) which can again be influenced by O-GlcNAcylation of Tet1 and regulation through Ogt (Vella et al., 2013). It could consequently be speculated that if H3K27me3 had already been deposited on some lineage-regulating genes in iPS cells, the effect of O-GlcNAc deregulation (decrease in Ogt protein levels) would be restricted to certain genes that would not carry this methylation mark. Hence, this could explain why a lineage-restricted effect can be observed in iPS cells, but a lineage-independent effect

in ES cells. However, the most probable cause for the observed differences between mouse ES cells and human iPS cells is a combination of small effects of GNS G treatment on multiple transcriptional regulatory mechanisms. As those mechanisms are in a different starting position when the cells are subjected to treatment and differentiation, the outcome differs between the two model systems. Ideally, experiments should have been repeated using both mouse iPS cells and human ES cells in order to provided insight into the comparability of both species and pluripotent stem cell model systems. In addition, differentiation experiments using mouse EpiSCs which are in developmental state that is thought to be similar to the one of human iPS cells (primed pluripotency) could be performed. Due to time restrictions such experiments have not been subject of the thesis at hand and are therefore strongly recommended for future work on the project.

6.4. O-GlcNAc in epigenetics

6.4.1. Reexpression of 2C genes and implications for cell potency

Experiments on mouse ES cells presented in the thesis at hand provide proof of reexpression genes that is normally silenced when treated with GNS G section 4.2. These genes have previously been described to be characteristic for the early developmental stage of the two-cell embryo (Macfarlan et al., 2012). Expression of 2C genes is correlated with retroviral activity as their transcription is initiated from long terminal repeats (LTRs) of murine endogenous retroviruses. As the embryo progresses in its development those retroviral elements and respective transcripts become silenced

through epigenetic mechanism. Thus, expression of 2C genes can be seen as an indicator of this early developmental two-cell stage at which cells are still totipotent. Macfarlan et al. (2012) suggest, that within every mouse ES cell population a totipotent subset exists that can be identified by expression of 2C genes. Furthermore, they suggest that ES cells cycle in and out of the totipotent state. Data presented in the thesis at hand clearly indicates that treatment with GNS G as well as knockdown of Oga increases transcription of 2C genes. Flow-cytometry experiments using MERVL reporter cells have demonstrated that the amplification in expression of 2C genes observed when investigating the bulk population is due to a increase of MERVL-expressing subpopulation. Therefore, GNS G treatment leads to an enlargement of a potentially totipotent subpopulation. In order to elucidate whether MERVL-expressing cells are actually a cycling population as suggested by Macfarlan et al. (2012) several attempts were made to sort the cells. Unfortunately, cell viability was greatly impaired in response to the sorting. For this reason the results are not included in this thesis. However, data obtained with the few surviving cells suggested that the MERVL subpopulation does cycle.

It remains to be clarified an increase in the number of cells that express 2C genes can be correlated with an effect on total cell potency. If so, this would indicate that GNS G treatment or knockdown of Oga could shift the cells towards a totipotent state. Macfarlan et al. (2012) have used morula injection of the sorted 2C-subpopulation in order to validate that those cells are able to contribute to the formation of extraembryonic tissue. Due to the above mentioned technical restriction, similar experiments have not been conducted as part of this thesis. Nonetheless, preliminary experiments have been carried aiming to differentiate GNS G treated mouse ES cells into trophoblast. However, results from those experiments were inconclusive. Furthermore,

we tested whether cells treated with GNS G would also show an increased expression of *Hex*, an extraembryonic endoderm marker that had been linked with totipotency (Morgani et al., 2013). Preliminary data did not indicate an increase in *Hex* expression upon GNS G treatment. Consequently, it remains elusive whether deregulation of O-GlcNAcylation can be linked with totipotency. In the light of data obtained to date, it could be speculated that GNS G treatment exhibits a broad de-repressive function on an epigenetic level that happens to include 2C genes amongst other gene sets, rather than providing a switch from a pluripotent to a totipotent state.

6.4.2. Polycomb repression

The previously presented results indicated that in addition to 2C genes a second gene set becomes re-expressed in mouse ES cells in response to GNS G treatment. This set is composed of genes that are normally silenced by polycomb repressive complexes (Speakman et al., 2014; Brookes et al., 2012). It should be noted that detailed investigations on the function of O-GlcNAc in polycomb repression were not subject of this thesis at any point. Nevertheless, the validation of gene set enrichment data presented by Speakman et al. (2014) can provide a starting point for future investigations. The following discussion will therefore aim to reveal potential links between the observations provided by the study at hand regarding pluripotent stem cell differentiation and gene regulation in order to stimulate further research. As described in section 1.2.2 O-GlcNAc and its regulating enzymes have been shown to crosstalk with polycomb repression in multiple studies. Nonetheless, it remains largely elusive to which extent O-GlcNAcylation affects the two complexes and how this could impact on differentiation and pluripotency. A very recent study provides insight into

potential mechanisms by which O-GlcNAc modification could regulate the activity of PRC1: In 2015, Maury et al. (2015) used human ES cells to provide evidence that the catalytical subunit of PRC1, Ring1B, can be modified by O-GlcNAc. Interestingly, the modification could be shown to correlate with localization of Ring1B on chromatin as the O-GlcNAcylated enzyme is preferentially found on genes that are related to the neural differentiation process. In addition, O-GlcNAc levels appear to decrease during differentiation. It is important to note that this observation is not entirely consistent with the data presented in this thesis. The observed discrepancies might be due to methodological differences. However, if O-GlcNAcylation of Ring1B decreases with the progress of differentiation, a loss of the modification could provide an important on-switch of neural genes. Therefore, GNS G treatment could artificially stabilize O-GlcNAcylation of Ring1B and might thereby hinder neural differentiation as relevant genes remain in a silenced state. Of note, the data presented by Maury et al. (2015) was obtained using human ES cells. As part of this thesis, transcriptome analysis of human iPS cells treated with GNS G was performed and this data does not suggest a significant deregulation of polycomb repression. Moreover, gene set enrichment analysis of microarray data on mouse ES cells performed in Speakman et al. (2014) indicate no significant enrichment for two additional gene sets that consist of target genes of PRC1 or PRC2. Myers et al. (2011) also demonstrate that knockdown of Ogt, which corresponds with decreasing levels of O-GlcNAcylation, does not impair PRC2 function. It is therefore unclear how pronounced the reported upregulation of polycomb repressed genes is and to what extent it impacts on cellular signalling. Potential effects could be limited to certain genes or loci and biological consequences would then be difficult to predict. An additional possibility is that the observed de-repression of polycomb targets in GNS G is a secondary effect that follows other gene

regulatory mechanisms, i.e. histone modification or DNA methylation. As an example, recruitment of PRC2 requires preceding demethylation of CpG islands which can be mediated by Tet1 which has been shown to interact with Ogt (Wu and Zhang, 2011; Aloia et al., 2013; Vella et al., 2013).

6.4.3. Histone acetylation

The previously described impact of O-GlcNAcylation on transcriptional regulation of 2C and polycomb repressive complex repressed genes might suggest an underlying more global mechanism. One candidate mechanism that has been investigated as a part of this thesis is a potential effect of O-GlcNAc on histone acetylation. An increase in histone acetylation is commonly understood to lead to a more open chromatin conformation, hence elevating transcriptional activity. O-GlcNAc is part of the histone code and therefore capable of interfering with respective histone modifications (Sakabe et al., 2010; Zhang et al., 2011; Fujiki et al., 2011; Fong et al., 2012). More importantly, there have been reports that the C-terminus of Oga harbours a domain that exhibits HAT activity (Schultz, 2002; Toleman et al., 2004). However, it has also been argued that this domain is catalytically inactive Rao et al. (2013); He et al. (2014). Thus, its function is still under debate. In the present study global histone acetylation was investigated using various methods. Initial experiments indicated an increase of lysine acetylation of H3 when investigated by Western Blot of histone extracts. However, strong variations were observed between individual experiments and overall reproducibility was low. Detailed records of cell morphology, density and viability for each experiments were kept, but did not help to elucidate these issues. As previously discussed, GNS G treatment has multiple effects and the feedback upregulation of

Oga protein levels is only one of them. Therefore, we decided to set up an experiment using knockdown and overexpression of Oga. Again, high variability was observed and the data was difficult to interpret. As a consequence, a different method was chosen for investigating H3 acetylation levels: A commercial ELISA kit was used on histone extracts of cells that had been transfected with overexpression plasmids of full-length Oga and a naturally occurring splicing-variant devoid of the major part of the HAT domain (Oga-SV). Independent experiments were performed and again records were kept on the cells' density and health status. There appears to be a minor trend towards an increase in H3 acetylation when cells were transfected with full-length Oga. However, results were again highly variable. It was observed, that the signal intensity of the colorimetric ELISA readout varied strongly with different batches of the kit regardless of the fact that consistent amounts of histone extracts were used. These observed batch effects make it even more difficult to confidently interpret the results obtained. Nonetheless, it appears that overexpression of Oga does not strongly impact on global H3 acetylation. Therefore, our results support the findings presented by Rao et al. (2013) and He et al. (2014) which indicate that Oga's HAT domain is catalytically inactive and can consequently be termed a pseudo-HAT domain. Regardless of the findings that Oga might not act as a HAT itself it might be speculated that the protein is involved in the regulation of histone modification. For example, the pseudo-HAT domain it might have a function in targeting other enzymes to histones or mediate complex formation of histone modifiers. If Oga acts on histones through interaction with other proteins it would be possible that upregulation of Oga alone is not sufficient to induce a effect that would be recognized by the methods applied in this study. Moreover, it should be noted that mouse ES cells represent a heterogeneous cell population and it has been proposed that a subpopulation with

extended cell potency exists. Therefore, a potential effect of Oga upregulation might be restricted to a certain subset within the mouse ES cell population. Consequently, effects on a small subpopulation may be partially masked when analysing the mouse ES cell bulk.

6.4.4. DNA methylation

Even though numerous previous studies indicate crosstalk between O-GlcNAcylation and Tet proteins which regulate DNA demethylation, the results presented in the thesis at hand do not suggest a major role of O-GlcNAc as a regulator of DNA methylation. Potential effects on DNA methylation were investigated using multiple approaches. Initial experiments were performed using bisulfite sequencing of CpG islands on gene bodies of selected genes and the respective results indicate no changes in DNA methylation of the investigated regions when cells were treated with GNS G. However, it can not be ruled out that the pattern of DNA methylation is modified on specific regulatory elements or genes that were not included in our analysis. In order to rule out an impact on few selected regions it would be recommended to perform a whole-genome bisulfite sequencing experiment which was beyond the scope of the present study. Instead, further experiments were performed investigating global DNA methylation using a dot blot method. Unfortunately, several technical issues impaired the interpretability of these experiments. The dot blot method appear to be very prone to error as the signal was often very low and therefore difficult to detect and interpret confidently. Moreover, results appeared to be highly variable, even within the technical dilutions performed for each experiment. Re-blotting of previously investigated DNA often lead to a different outcome that could not be explained by differences in

DNA storage or processing. Nonetheless, several consecutive repeats could be performed successfully and the results indicated an effect of GNS G treatment on DNA methylation. As we had little confidence in the dot blot method itself, we decided to perform a highly quantitative HPLC-MS experiment. Strikingly, results from this investigation very clearly indicate that levels of 5-mC and 5-hmC are not affected by GNS G treatment in mouse ES or human iPS cells.

These results are particularly surprising the in light of current knowledge on interactions of O-GlcNAc with Tet proteins. All three Tet proteins have been shown to be O-GlcNAc targets (Deplus et al., 2013; Ito et al., 2013; Shi et al., 2013) and at least Tet1 and Tet3 are regulated by the modification (Shi et al., 2013; Zhang et al., 2014). Moreover, Ogt interacts with both Tet1 and Tet3 and a role for this interaction in formation of repressive complexes has been proposed (Vella et al., 2013). GNS G treatment leads to a downregulation of Ogt protein levels. If Tet1/Ogt interact to recruit repressive complexes to chromatin, a loss of Ogt could correspond to a loss of repressive function. Similarly, Ogt promotes the nuclear export of Tet3 and thereby hinders conversion from 5-mC to 5-hmC (= demethylation) (Zhang et al., 2014). Less Ogt would allow more nuclear Tet3 which mediates demethylation and consequently leads to a loss of gene repression by DNA methylation. Both scenarios would correspond with the previously presented data on re-expression of silence gene sets. Intriguingly, our investigations of DNA methylation by bisulfite sequencing, dot blots and HPLC-MS do not support these theories. It could still be speculated that GNS G treatment has effects on DNA methylation that are restricted to single genes or selected regulatory elements. Therefore, whole-genome bisulfite sequencing might be a possibility to finally determine whether the crosstalk between O-GlcNAc and Tet proteins has any functional relevance regarding DNA methylation.

6.5. Implications from transcriptome analysis

Results obtained from transcriptome analysis using RNA sequencing have provided several important insights into the transcriptional changes in response to GNS G treatment. The key observation from this experiment is that O-GlcNAc appears to be globally involved in multiple cellular processes as a large number of transcripts with various functions are differentially expressed. In addition, GNS G treatment appears to have relatively small, yet significant effects on differentially expressed transcripts, meaning many genes are up- or downregulated, but only to a small extent. Consistent with our observations in mouse ES cells, there are more genes up- than downregulated which suggests a derepressive effect also in human iPS cells (Speakman et al., 2014). These observations underline that O-GlcNAc is a global player within the cell's signalling network that orchestrates multiple processes through precise transcriptional regulation. This "fine-tuning" function of O-GlcNAc might also be the reason why we could not determine a function of O-GlcNAc regarding selected transcriptional regulation mechanisms, such as histone acetylation or DNA methylation. The effects of deregulated O-GlcNAcylation could possibly be too small and coordinated very closely with others that investigations of single mechanisms might not allow for confident conclusions to be drawn. Despite the fact that this global effect makes it difficult to make predictions regarding a molecular mechanism that explains the impairment of neural differentiation, results obtained by gene ontology analysis provide validation current knowledge and reveal potential areas important for future research. For example, the analysis' results validate the inhibitor treatment as well as more general functions of O-GlcNAc by impacting on transcripts linked with phosphorylation and nutrition. More importantly, the analysis indicates a strong role of O-GlcNAc regarding neural

development. In the light of the previously obtained data on ectoderm differentiation this may appear little surprising. However, the deregulation of transcripts that are involved in processes like neural tube and brain development, neurogenesis and axon guidance strongly suggest a function of O-GlcNAcylation that stretches beyond stem cell differentiation. It would be most valuable to uncover the functions of O-GlcNAc in the development of the mammalian central nervous system. As O-GlcNAc is a key nutrient sensor, it may even provide a link to how the nutritional status of an organism impacts on its neural development or, through epigenetic imprints, even on that of its offspring.

Even though the results indicate multiple global mechanisms through which O-GlcNAc may impact on neural differentiation it was possible to identify a potential signalling pathway that appears to be affected by GNS G treatment. Analysis of differential gene expression identified a set of genes involved in BMP signalling to be differentially regulated in GNS G. Preliminary data suggests that human iPS cells which have been treated with GNS G exhibit increased sensitivity towards BMP4 stimulation. BMP signalling is part of the TGF-β cascade and plays a key role in the regulation of pluripotency and differentiation. In human ES cells, BMP signalling is blocked to maintain a pluripotent state and stimulation of the pathway is correlated with initiation of differentiation. However, in our experiments ectodermal differentiation was hindered when cells were treated with GNS G. Therefore, it remains unclear how increased sensitivity towards BMP signalling blocks differentiation rather than driving it. Interestingly, there are multiple studies underlining the importance of BMP4 signalling during embryogenesis of various vertebrate model systems. For example, BMP4 acts as a morphogen and is essential for dorso-ventral patterning in *Xenopus laevis* (Wilson et al., 1997; Dosch et al., 1997). Importantly, Khokha et al.

(2005) showed that inhibition of BMP signalling is required for formation of the neural plate in *Xenopus tropicalis*. Using zebrafish Koshida et al. (2002) demonstrated that blockage of BMP4 activity through FGF promotes neurogenesis. LaVaute et al. (2009) extended these findings to human ES cells: Their study confirms that inhibition of BMP signalling is necessary for neural differentiation. The data presented by LaVaute et al. (2009) support my findings in human iPS cells treated with GNS G: Inhibition of BMP signalling as required for neural specification might diminished due to an increase in the cells' sensitivity towards BMP stimulation. This hypothesis would also explain why the cells are capable of exiting the pluripotent state but unable to attain a neural identity. In addition, the previously presented data also indicates that GNS G no longer blocks neural differentiation when added at a later timepoint. The impact of BMP signalling might vary depending on how far the differentiation has progressed. Moreover, it is also likely that the cells require a certain period of time in GNS G to acquire an elevated sensitivity to BMP4. It is clear that detailed investigations on the role of O-GlcNAc in BMP/TGF-β signalling during differentiation are needed for a complete understanding this mechanism.

6.6. Future directions

The project at hand provides many important insights into the function of O-GlcNAc in pluripotent stem cells. Yet, many questions remain and provide various potential directions for further research on this topic. Results presented in this thesis clearly indicate discrepancies between the two pluripotent stem cell models used. It remains unclear to what extent those differences are species-specific or model-specific. Therefore, it would be highly recommended to perform additional experiments on differen-

tiation of human ES cells and mouse iPS cells to provide clarification on this matter. Most experiments performed for this study used GNS G as Oga inhibitor to elevate global O-GlcNAc levels. As previously discussed, GNS G treatment has the additional feature of inducing feedback regulatory effects on the two O-GlcNAc regulators, Oga and Ogt. These additional feedback effects make it difficult to determine molecular mechanisms as both Oga and Ogt have been implied with additional functions on the level of gene regulation. Therefore, it is essential to conduct a set of experiments using knockdown or overexpression of the two enzymes, ideally through an inducible and stable system. The generation of respective lines has been attempted, but was so far unsuccessful. Some of the difficulties of deregulating levels of Ogt and Oga are most likely due to the fact that the O-GlcNAc system is tightly regulated and disruptions are in many cases not tolerated by the cell. It might be worthwhile to apply a genetic engineering approach in order to generate ES/iPS cells lines with a monoallelic deletion which could even be limited to the catalytic domain if necessary.

Ideally, an additional set of RNA sequencing experiments using Ogt/Oga knockdown or overexpression lines could be performed to provide an understanding of the molecular reasons behind some of the effects in GNS G treated human iPS cells. If further transcriptome analysis would be performed, it might also prove insightful to investigate mouse and human pluripotent cells in direct comparison, ideally using mouse and human ES cells rather than iPS cells. Our lab has previously performed microarray-based analysis of gene expression of mouse ES cells in GNS C. Unfortunately, RNA sequencing and microarray analysis provide approaches for transcriptome analysis that are highly divergent on a technical level and therefore not suitable for direct comparison. Gene ontology of RNA sequencing data in GNS G treated human iPS cells revealed that many genes which are directly associated with development of

the central nervous system are affected by GNS G treatment. A potential approach to investigate further would be to study the functions of O-GlcNAc in neural development using a mammalian model organism like the mouse. Such investigations would obviously require the laborious generation of respective mouse lines but could still be worthwhile as several very recent studies imply an essential role of O-GlcNAc in the brain (Lagerlof et al., 2016; Gong et al., 2016).

Finally, additional experiments should be undertaken to build on the preliminary data suggesting a function of O-GlcNAc in the regulation of BMP signalling. First of all, the preliminary data that indicates an elevated sensitivity of GNS G treated cells towards BMP4 stimulation should be repeated. Additional experiments could then focus on revealing the consequences on the entire signalling pathway, but also investigate a potential effect on the whole TGF-β cascade. When studying the impact on BMP signalling it is essential to perform these experiments both in pluripotent cells and differentiating cells, ideally using a detailed timecourse of neural differentiation. In addition, the question arises whether key components of the BMP signalling pathway are modified and regulated by O-GlcNAc in human iPS cells. A proteomic approach could provide the respective information. It would also be interesting to investigate how the landscape of O-GlcNAcylated proteins changes over the course of differentiation and to what extent removal or addition of the modifications affects the cell's ability to differentiate.

A. Appendix

A.1. Supplementary RNA sequencing data

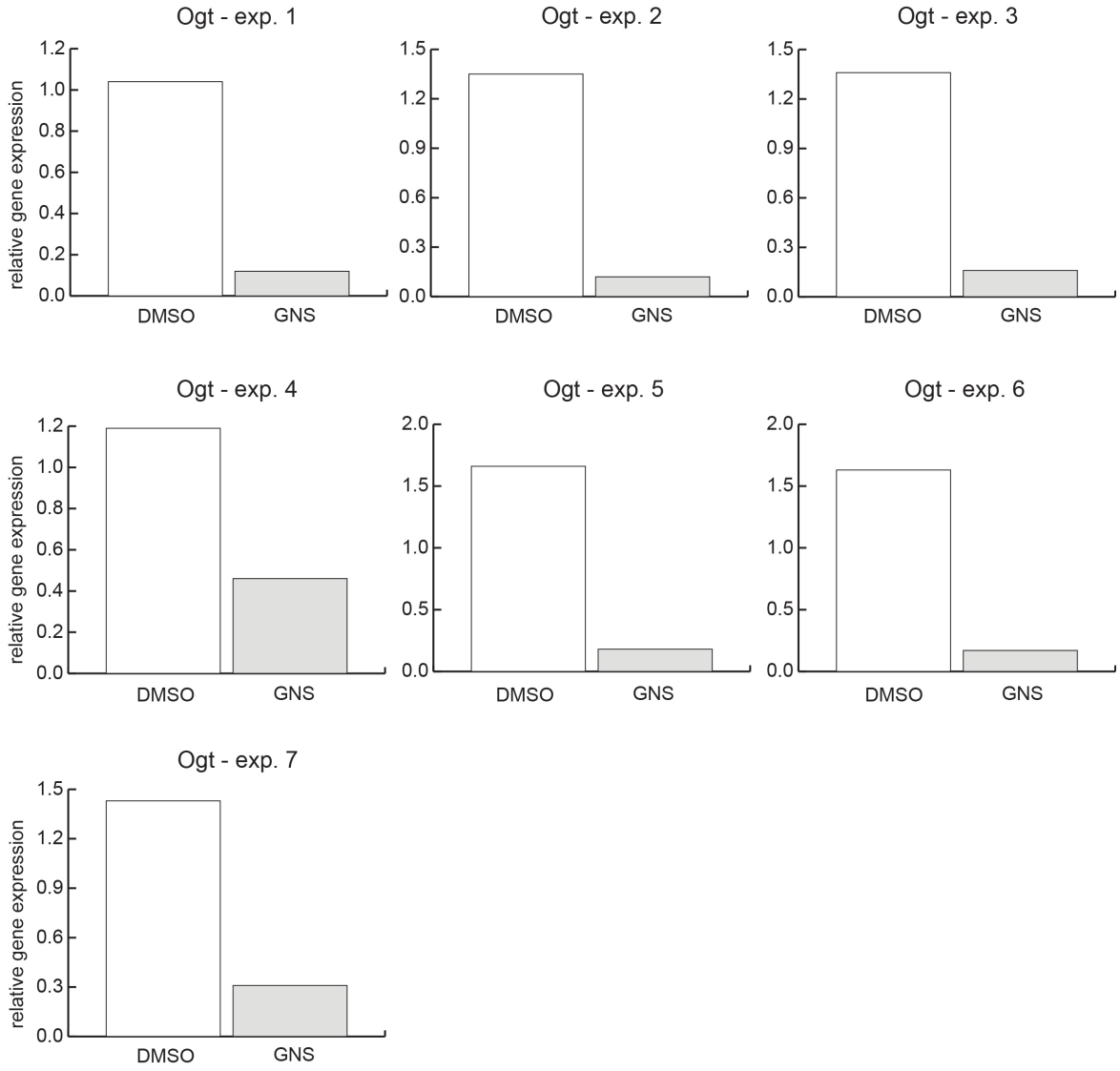

Figure (A.1) RT-qPCR analysis of RNA sequencing samples for Ogt expression.

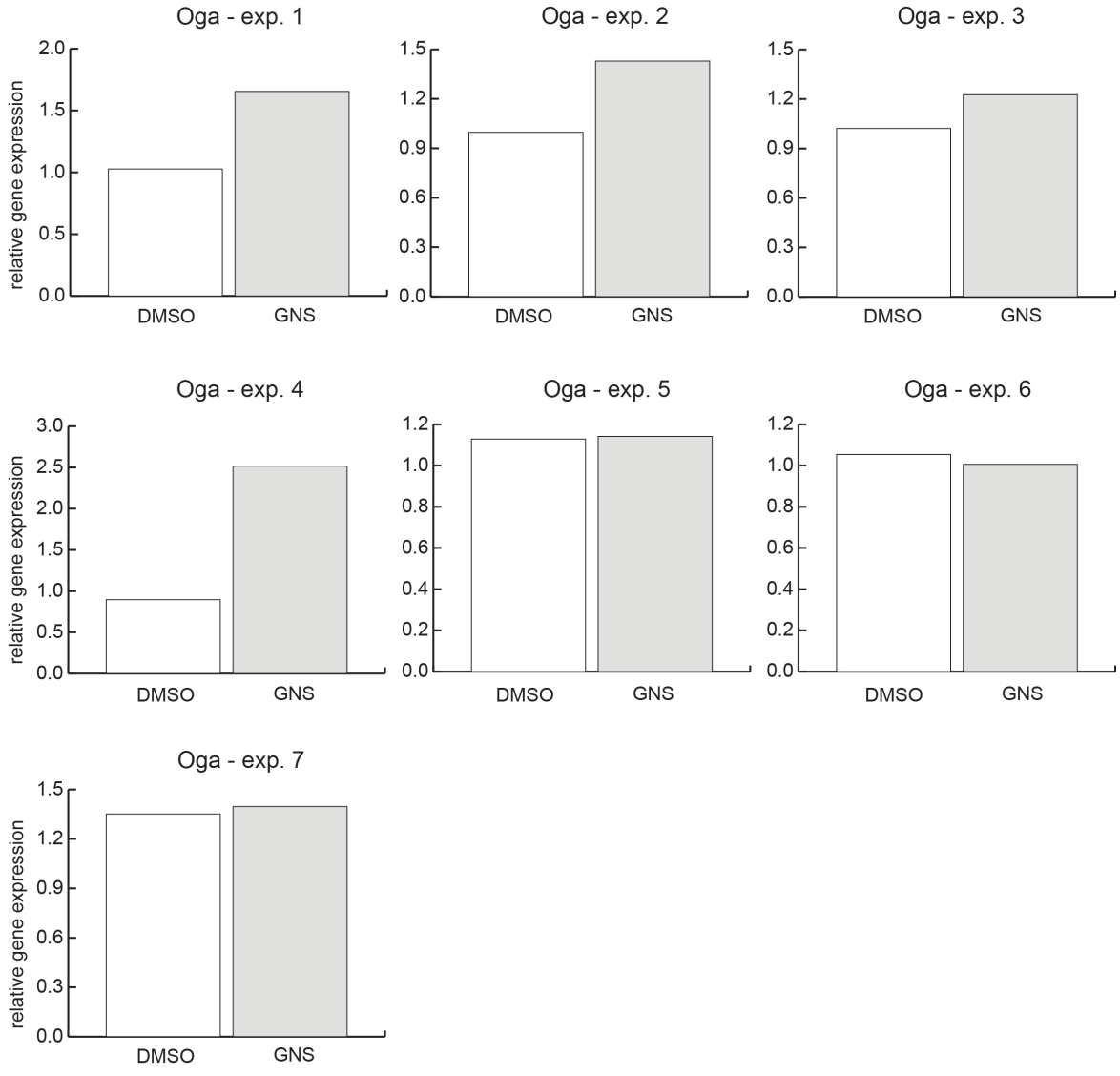

Figure (A.2) RT-qPCR analysis of RNA sequencing samples for Oga expression.

A.1.1. Differential gene expression with EdgeR

Table (A.1) Differentially expressed genes with EdgeR (logFC $\geq$ 0.25, FDR<0.01)

	ensembl_gene_id	external_gene_name	description	logFC	FDR
1	ENSG00000203900	RP11-261N11.8		-4.22	0.00
2	ENSG00000182870	GALNT9	polypeptide N-acetylgalactosaminyltransferase 9 [Source:HGNC Symbol;Acc:HGNC:4131]	-2.59	0.00
3	ENSG00000171246	NPTX1	neuronal pentraxin I [Source:HGNC Symbol;Acc:HGNC:7952]	-2.40	0.00
4	ENSG00000268869	ESPNP	espin pseudogene [Source:HGNC Symbol;Acc:HGNC:23285]	-1.92	0.00
5	ENSG00000127528	KLF2	Kruppel-like factor 2 [Source:HGNC Symbol;Acc:HGNC:6347]	-1.92	0.00
6	ENSG00000128610	FEZF1	FEZ family zinc finger 1 [Source:HGNC Symbol;Acc:HGNC:22788]	-1.91	0.00
7	ENSG00000260750	RP11-482M8.1		-1.90	0.00
8	ENSG00000185105	MYADML2	myeloid-associated differentiation marker-like 2 [Source:HGNC Symbol;Acc:HGNC:34548]	-1.79	0.00
9	ENSG00000089116	LHX5	LIM homeobox 5 [Source:HGNC Symbol;Acc:HGNC:14216]	-1.72	0.00
10	ENSG00000205116	TMEM88B	transmembrane protein 88B [Source:HGNC Symbol;Acc:HGNC:37099]	-1.71	0.00
11	ENSG00000060558	GNA15	guanine nucleotide binding protein (G protein), alpha 15 (Gq class) [Source:HGNC Symbol;Acc:HGNC:4383]	-1.69	0.00
12	ENSG00000196381	ZNF781	zinc finger protein 781 [Source:HGNC Symbol;Acc:HGNC:26745]	-1.65	0.00
13	ENSG00000230316	FEZF1-AS1	FEZF1 antisense RNA 1 [Source:HGNC Symbol;Acc:HGNC:41001]	-1.55	0.00
14	ENSG00000134594	RAB33A	RAB33A, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:9773]	-1.51	0.01
15	ENSG00000101222	SPEF1	sperm flagellar 1 [Source:HGNC Symbol;Acc:HGNC:15874]	-1.51	0.00
16	ENSG00000130643	CALY	calcyon neuron-specific vesicular protein [Source:HGNC Symbol;Acc:HGNC:17938]	-1.48	0.00
17	ENSG00000225783	MIAT	myocardial infarction associated transcript (non-protein coding) [Source:HGNC Symbol;Acc:HGNC:33425]	-1.47	0.00
18	ENSG00000141665	FBXO15	F-box protein 15 [Source:HGNC Symbol;Acc:HGNC:13617]	-1.44	0.00
19	ENSG00000205464	ATP6AP1L	ATPase, H+ transporting, lysosomal accessory protein 1-like [Source:HGNC Symbol;Acc:HGNC:28091]	-1.44	0.00
20	ENSG00000100314	CABP7	calcium binding protein 7 [Source:HGNC Symbol;Acc:HGNC:20834]	-1.42	0.00
21	ENSG00000179292	TMEM151A	transmembrane protein 151A [Source:HGNC Symbol;Acc:HGNC:28497]	-1.41	0.00
22	ENSG00000174326	SLC16A11	solute carrier family 16 member 11 [Source:HGNC Symbol;Acc:HGNC:23093]	-1.41	0.00
23	ENSG00000249550	LINC01234	long intergenic non-protein coding RNA 1234 [Source:HGNC Symbol;Acc:HGNC:49757]	-1.41	0.00
24	ENSG00000177875	CCDC184	coiled-coil domain containing 184 [Source:HGNC Symbol;Acc:HGNC:33749]	-1.39	0.00
25	ENSG00000146216	TTBK1	tau tubulin kinase 1 [Source:HGNC Symbol;Acc:HGNC:19140]	-1.37	0.00
26	ENSG00000152932	RAB3C	RAB3C, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:30269]	-1.36	0.00
27	ENSG00000188886	ASTL	astacin-like metallo-endopeptidase (M12 family) [Source:HGNC Symbol;Acc:HGNC:31704]	-1.34	0.00
28	ENSG00000008735	MAPK8IP2	mitogen-activated protein kinase 8 interacting protein 2 [Source:HGNC Symbol;Acc:HGNC:6883]	-1.32	0.00
29	ENSG00000069696	DRD4	dopamine receptor D4 [Source:HGNC Symbol;Acc:HGNC:3025]	-1.31	0.00
30	ENSG00000187902	SHISA7	shisa family member 7 [Source:HGNC Symbol;Acc:HGNC:35409]	-1.30	0.01
31	ENSG00000106018	VIPR2	vasoactive intestinal peptide receptor 2 [Source:HGNC Symbol;Acc:HGNC:12695]	-1.29	0.00
32	ENSG00000165495	PKNOX2	PBX/knotted 1 homeobox 2 [Source:HGNC Symbol;Acc:HGNC:16714]	-1.29	0.00
33	ENSG00000176182	MYPOP	Myb-related transcription factor, partner of profilin [Source:HGNC Symbol;Acc:HGNC:20178]	-1.28	0.00

34	ENSG00000186891	TNFRSF18	tumor necrosis factor receptor superfamily member 18 [Source:HGNC Symbol;Acc:HGNC:11914]	-1.28	0.01
35	ENSG00000177807	KCNJ10	potassium channel, inwardly rectifying subfamily J, member 10 [Source:HGNC Symbol;Acc:HGNC:6256]	-1.28	0.00
36	ENSG00000130427	EPO	erythropoietin [Source:HGNC Symbol;Acc:HGNC:3415]	-1.25	0.00
37	ENSG00000125388	GRK4	G protein-coupled receptor kinase 4 [Source:HGNC Symbol;Acc:HGNC:4543]	-1.25	0.00
38	ENSG00000138311	ZNF365	zinc finger protein 365 [Source:HGNC Symbol;Acc:HGNC:18194]	-1.24	0.00
39	ENSG00000132016	C19orf57	chromosome 19 open reading frame 57 [Source:HGNC Symbol;Acc:HGNC:28153]	-1.24	0.00
40	ENSG00000182566	CLEC4G	C-type lectin domain family 4 member G [Source:HGNC Symbol;Acc:HGNC:24591]	-1.21	0.00
41	ENSG00000159167	STC1	stanniocalcin 1 [Source:HGNC Symbol;Acc:HGNC:11373]	-1.21	0.00
42	ENSG00000125285	SOX21	SRY-box 21 [Source:HGNC Symbol;Acc:HGNC:11197]	-1.21	0.00
43	ENSG00000063015	SEZ6	seizure related 6 homolog (mouse) [Source:HGNC Symbol;Acc:HGNC:15955]	-1.20	0.00
44	ENSG00000078967	UBE2D4	ubiquitin conjugating enzyme E2D 4 (putative) [Source:HGNC Symbol;Acc:HGNC:21647]	-1.19	0.00
45	ENSG00000198720	ANKRD13B	ankyrin repeat domain 13B [Source:HGNC Symbol;Acc:HGNC:26363]	-1.18	0.00
46	ENSG00000161149	TUBA3FP	tubulin alpha 3f pseudogene [Source:HGNC Symbol;Acc:HGNC:24067]	-1.17	0.00
47	ENSG00000184524	CEND1	cell cycle exit and neuronal differentiation 1 [Source:HGNC Symbol;Acc:HGNC:24153]	-1.16	0.00
48	ENSG00000196169	KIF19	kinesin family member 19 [Source:HGNC Symbol;Acc:HGNC:26735]	-1.13	0.00
49	ENSG00000162639	HENMT1	HEN1 methyltransferase homolog 1 (Arabidopsis) [Source:HGNC Symbol;Acc:HGNC:26400]	-1.12	0.00
50	ENSG00000122641	INHBA	inhibin beta A [Source:HGNC Symbol;Acc:HGNC:6066]	-1.11	0.00
51	ENSG00000104332	SFRP1	secreted frizzled-related protein 1 [Source:HGNC Symbol;Acc:HGNC:10776]	-1.11	0.00
52	ENSG00000132141	CCT6B	chaperonin containing TCP1, subunit 6B (zeta 2) [Source:HGNC Symbol;Acc:HGNC:1621]	-1.11	0.00
53	ENSG00000102981	PARD6A	par-6 family cell polarity regulator alpha [Source:HGNC Symbol;Acc:HGNC:15943]	-1.08	0.00
54	ENSG00000249853	HS3ST5	heparan sulfate-glucosamine 3-sulfotransferase 5 [Source:HGNC Symbol;Acc:HGNC:19419]	-1.07	0.00
55	ENSG00000196155	PLEKHG4	pleckstrin homology and RhoGEF domain containing G4 [Source:HGNC Symbol;Acc:HGNC:24501]	-1.07	0.00
56	ENSG00000189410	SH2D5	SH2 domain containing 5 [Source:HGNC Symbol;Acc:HGNC:28819]	-1.06	0.00
57	ENSG00000138400	MDH1B	malate dehydrogenase 1B [Source:HGNC Symbol;Acc:HGNC:17836]	-1.06	0.00
58	ENSG00000159761	C16orf86	chromosome 16 open reading frame 86 [Source:HGNC Symbol;Acc:HGNC:33755]	-1.06	0.00
59	ENSG00000215644	GCGR	glucagon receptor [Source:HGNC Symbol;Acc:HGNC:4192]	-1.05	0.00
60	ENSG00000145882	PCYOX1L	prenylcysteine oxidase 1 like [Source:HGNC Symbol;Acc:HGNC:28477]	-1.05	0.00
61	ENSG00000184208	C22orf46	chromosome 22 open reading frame 46 [Source:HGNC Symbol;Acc:HGNC:26294]	-1.04	0.00
62	ENSG00000139793	MBNL2	muscleblind-like splicing regulator 2 [Source:HGNC Symbol;Acc:HGNC:16746]	-1.04	0.00
63	ENSG00000148204	CRB2	crumbs family member 2 [Source:HGNC Symbol;Acc:HGNC:18688]	-1.04	0.00
64	ENSG00000177181	RIMKLA	ribosomal modification protein rimK-like family member A [Source:HGNC Symbol;Acc:HGNC:28725]	-1.03	0.00
65	ENSG00000128872	TMOD2	tropomodulin 2 [Source:HGNC Symbol;Acc:HGNC:11872]	-1.03	0.00
66	ENSG00000157851	DPYSL5	dihydropyrimidinase like 5 [Source:HGNC Symbol;Acc:HGNC:20637]	-1.02	0.00
67	ENSG00000168453	HR	hair growth associated [Source:HGNC Symbol;Acc:HGNC:5172]	-1.02	0.00
68	ENSG00000143643	TTC13	tetratricopeptide repeat domain 13 [Source:HGNC Symbol;Acc:HGNC:26204]	-1.02	0.00
69	ENSG00000162426	SLC45A1	solute carrier family 45 member 1 [Source:HGNC Symbol;Acc:HGNC:17939]	-1.01	0.00

70	ENSG00000021300	PLEKHB1	pleckstrin homology domain containing B1 [Source:HGNC Symbol;Acc:HGNC:19079]	-1.00	0.00
71	ENSG00000025708	TYMP	thymidine phosphorylase [Source:HGNC Symbol;Acc:HGNC:3148]	-0.99	0.00
72	ENSG00000105376	ICAM5	intercellular adhesion molecule 5 [Source:HGNC Symbol;Acc:HGNC:5348]	-0.99	0.00
73	ENSG00000152804	HHEX	hematopoietically expressed homeobox [Source:HGNC Symbol;Acc:HGNC:4901]	-0.98	0.00
74	ENSG00000204934	ATP6V0E2-AS1	ATP6V0E2 antisense RNA 1 [Source:HGNC Symbol;Acc:HGNC:44180]	-0.98	0.00
75	ENSG00000185561	TLCD2	TLC domain containing 2 [Source:HGNC Symbol;Acc:HGNC:33522]	-0.97	0.00
76	ENSG00000095752	IL11	interleukin 11 [Source:HGNC Symbol;Acc:HGNC:5966]	-0.97	0.00
77	ENSG00000102003	SYP	synaptophysin [Source:HGNC Symbol;Acc:HGNC:11506]	-0.97	0.00
78	ENSG00000182134	TDRKH	tudor and KH domain containing [Source:HGNC Symbol;Acc:HGNC:11713]	-0.97	0.00
79	ENSG00000068976	PYGM	phosphorylase, glycogen, muscle [Source:HGNC Symbol;Acc:HGNC:9726]	-0.97	0.00
80	ENSG00000171533	MAP6	microtubule associated protein 6 [Source:HGNC Symbol;Acc:HGNC:6868]	-0.96	0.00
81	ENSG00000132436	FIGNL1	fidgetin-like 1 [Source:HGNC Symbol;Acc:HGNC:13286]	-0.96	0.00
82	ENSG00000212978	AC016747.3		-0.96	0.00
83	ENSG00000183773	AIFM3	apoptosis inducing factor, mitochondria associated 3 [Source:HGNC Symbol;Acc:HGNC:26398]	-0.95	0.01
84	ENSG00000215271	HOMEZ	homeobox and leucine zipper encoding [Source:HGNC Symbol;Acc:HGNC:20164]	-0.95	0.00
85	ENSG00000187583	PLEKHN1	pleckstrin homology domain containing N1 [Source:HGNC Symbol;Acc:HGNC:25284]	-0.95	0.01
86	ENSG00000121454	LHX4	LIM homeobox 4 [Source:HGNC Symbol;Acc:HGNC:21734]	-0.94	0.00
87	ENSG00000112561	TFEB	transcription factor EB [Source:HGNC Symbol;Acc:HGNC:11753]	-0.94	0.00
88	ENSG00000090530	P3H2	prolyl 3-hydroxylase 2 [Source:HGNC Symbol;Acc:HGNC:19317]	-0.94	0.00
89	ENSG00000099282	TSPAN15	tetraspanin 15 [Source:HGNC Symbol;Acc:HGNC:23298]	-0.94	0.00
90	ENSG00000178233	TMEM151B	transmembrane protein 151B [Source:HGNC Symbol;Acc:HGNC:21315]	-0.93	0.00
91	ENSG00000162572	SCNN1D	sodium channel, non voltage gated 1 delta subunit [Source:HGNC Symbol;Acc:HGNC:10601]	-0.93	0.00
92	ENSG00000159733	ZFYVE28	zinc finger, FYVE domain containing 28 [Source:HGNC Symbol;Acc:HGNC:29334]	-0.92	0.00
93	ENSG00000185269	NOTUM	notum pectinacetylesterase homolog (Drosophila) [Source:HGNC Symbol;Acc:HGNC:27106]	-0.92	0.00
94	ENSG00000105251	SHD	Src homology 2 domain containing transforming protein D [Source:HGNC Symbol;Acc:HGNC:30633]	-0.90	0.00
95	ENSG00000180425	C11orf71	chromosome 11 open reading frame 71 [Source:HGNC Symbol;Acc:HGNC:25937]	-0.90	0.00
96	ENSG00000177570	SAMD12	sterile alpha motif domain containing 12 [Source:HGNC Symbol;Acc:HGNC:31750]	-0.90	0.00
97	ENSG00000149599	DUSP15	dual specificity phosphatase 15 [Source:HGNC Symbol;Acc:HGNC:16236]	-0.90	0.00
98	ENSG00000163617	CCDC191	coiled-coil domain containing 191 [Source:HGNC Symbol;Acc:HGNC:29272]	-0.90	0.00
99	ENSG00000185386	MAPK11	mitogen-activated protein kinase 11 [Source:HGNC Symbol;Acc:HGNC:6873]	-0.90	0.00
100	ENSG00000139800	ZIC5	Zic family member 5 [Source:HGNC Symbol;Acc:HGNC:20322]	-0.89	0.00
101	ENSG00000273045	C2orf15	chromosome 2 open reading frame 15 [Source:HGNC Symbol;Acc:HGNC:28436]	-0.89	0.00
102	ENSG00000125945	ZNF436	zinc finger protein 436 [Source:HGNC Symbol;Acc:HGNC:20814]	-0.89	0.00
103	ENSG00000228594	C1orf233	chromosome 1 open reading frame 233 [Source:HGNC Symbol;Acc:HGNC:42951]	-0.89	0.00
104	ENSG00000175414	ARL10	ADP ribosylation factor like GTPase 10 [Source:HGNC Symbol;Acc:HGNC:22042]	-0.88	0.00
105	ENSG00000113648	H2AFY	H2A histone family member Y [Source:HGNC Symbol;Acc:HGNC:4740]	-0.88	0.00
106	ENSG00000228889	UBAC2-AS1	UBAC2 antisense RNA 1 [Source:HGNC Symbol;Acc:HGNC:42502]	-0.88	0.00
107	ENSG00000185621	LMLN	leishmanolysin-like (metallopeptidase M8 family) [Source:HGNC Symbol;Acc:HGNC:15991]	-0.88	0.00
108	ENSG00000101417	PXMP4	peroxisomal membrane protein 4 [Source:HGNC Symbol;Acc:HGNC:15920]	-0.88	0.00
109	ENSG00000109832	DDX25	DEAD (Asp-Glu-Ala-Asp) box helicase 25 [Source:HGNC Symbol;Acc:HGNC:18698]	-0.88	0.00

110	ENSG00000164885	CDK5	cyclin-dependent kinase 5 [Source:HGNC Symbol;Acc:HGNC:1774]	-0.88	0.00
111	ENSG00000198208	RPS6KL1	ribosomal protein S6 kinase like 1 [Source:HGNC Symbol;Acc:HGNC:20222]	-0.87	0.00
112	ENSG00000168490	PHYHIP	phytanoyl-CoA 2-hydroxylase interacting protein [Source:HGNC Symbol;Acc:HGNC:16865]	-0.87	0.00
113	ENSG00000107014	RLN2	relaxin 2 [Source:HGNC Symbol;Acc:HGNC:10027]	-0.87	0.00
114	ENSG00000129654	FOXJ1	forkhead box J1 [Source:HGNC Symbol;Acc:HGNC:3816]	-0.87	0.00
115	ENSG00000102109	PCSK1N	proprotein convertase subtilisin/kexin type 1 inhibitor [Source:HGNC Symbol;Acc:HGNC:17301]	-0.86	0.00
116	ENSG00000136274	NACAD	NAC alpha domain containing [Source:HGNC Symbol;Acc:HGNC:22196]	-0.86	0.00
117	ENSG00000078487	ZCWPW1	zinc finger, CW type with PWWP domain 1 [Source:HGNC Symbol;Acc:HGNC:23486]	-0.86	0.00
118	ENSG00000133794	ARNTL	aryl hydrocarbon receptor nuclear translocator like [Source:HGNC Symbol;Acc:HGNC:701]	-0.85	0.00
119	ENSG00000118513	MYB	v-myb avian myeloblastosis viral oncogene homolog [Source:HGNC Symbol;Acc:HGNC:7545]	-0.84	0.00
120	ENSG00000079974	RABL2B	RAB, member of RAS oncogene family-like 2B [Source:HGNC Symbol;Acc:HGNC:9800]	-0.84	0.00
121	ENSG00000148926	ADM	adrenomedullin [Source:HGNC Symbol;Acc:HGNC:259]	-0.83	0.00
122	ENSG00000185567	AHNAK2	AHNAK nucleoprotein 2 [Source:HGNC Symbol;Acc:HGNC:20125]	-0.83	0.00
123	ENSG00000241360	PDXP	pyridoxal (pyridoxine, vitamin B6) phosphatase [Source:HGNC Symbol;Acc:HGNC:30259]	-0.83	0.00
124	ENSG00000213963	AC074286.1		-0.83	0.00
125	ENSG00000043355	ZIC2	Zic family member 2 [Source:HGNC Symbol;Acc:HGNC:12873]	-0.83	0.00
126	ENSG00000130287	NCAN	neurocan [Source:HGNC Symbol;Acc:HGNC:2465]	-0.83	0.00
127	ENSG00000160460	SPTBN4	spectrin, beta, non-erythrocytic 4 [Source:HGNC Symbol;Acc:HGNC:14896]	-0.83	0.00
128	ENSG00000167711	SERPINF2	serpin peptidase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 2 [Source:HGNC Symbol;Acc:HGNC:9075]	-0.82	0.00
129	ENSG00000154930	ACSS1	acyl-CoA synthetase short-chain family member 1 [Source:HGNC Symbol;Acc:HGNC:16091]	-0.82	0.00
130	ENSG00000166796	LDHC	lactate dehydrogenase C [Source:HGNC Symbol;Acc:HGNC:6544]	-0.81	0.00
131	ENSG00000245498	RP11-677M14.7		-0.81	0.00
132	ENSG00000173175	ADCY5	adenylate cyclase 5 [Source:HGNC Symbol;Acc:HGNC:236]	-0.80	0.00
133	ENSG00000106013	ANKRD7	ankyrin repeat domain 7 [Source:HGNC Symbol;Acc:HGNC:18588]	-0.80	0.00
134	ENSG00000166780	C16orf45	chromosome 16 open reading frame 45 [Source:HGNC Symbol;Acc:HGNC:19213]	-0.80	0.00
135	ENSG00000185761	ADAMTSL5	ADAMTS like 5 [Source:HGNC Symbol;Acc:HGNC:27912]	-0.80	0.00
136	ENSG00000132881	RSG1	REM2 and RAB-like small GTPase 1 [Source:HGNC Symbol;Acc:HGNC:28127]	-0.79	0.00
137	ENSG00000243449	C4orf48	chromosome 4 open reading frame 48 [Source:HGNC Symbol;Acc:HGNC:34437]	-0.79	0.00
138	ENSG00000170231	FABP6	fatty acid binding protein 6, ileal [Source:HGNC Symbol;Acc:HGNC:3561]	-0.79	0.00
139	ENSG00000133640	LRRIQ1	leucine-rich repeats and IQ motif containing 1 [Source:HGNC Symbol;Acc:HGNC:25708]	-0.79	0.00
140	ENSG00000156253	RWDD2B	RWD domain containing 2B [Source:HGNC Symbol;Acc:HGNC:1302]	-0.79	0.00
141	ENSG00000271601	LIX1L	limb and CNS expressed 1 like [Source:HGNC Symbol;Acc:HGNC:28715]	-0.78	0.00
142	ENSG00000041880	PARP3	poly(ADP-ribose) polymerase family member 3 [Source:HGNC Symbol;Acc:HGNC:273]	-0.78	0.00
143	ENSG00000160471	COX6B2	cytochrome c oxidase subunit VIb polypeptide 2 (testis) [Source:HGNC Symbol;Acc:HGNC:24380]	-0.78	0.00

144	ENSG00000139266	MARCH9	membrane associated ring-CH-type finger 9 [Source:HGNC Symbol;Acc:HGNC:25139]	-0.78	0.00
145	ENSG00000169258	GPRIN1	G protein regulated inducer of neurite outgrowth 1 [Source:HGNC Symbol;Acc:HGNC:24835]	-0.78	0.00
146	ENSG00000147117	ZNF157	zinc finger protein 157 [Source:HGNC Symbol;Acc:HGNC:12942]	-0.78	0.00
147	ENSG00000161860	SYCE2	synaptonemal complex central element protein 2 [Source:HGNC Symbol;Acc:HGNC:27411]	-0.77	0.00
148	ENSG00000088876	ZNF343	zinc finger protein 343 [Source:HGNC Symbol;Acc:HGNC:16017]	-0.77	0.00
149	ENSG00000149548	CCDC15	coiled-coil domain containing 15 [Source:HGNC Symbol;Acc:HGNC:25798]	-0.77	0.00
150	ENSG00000183570	PCBP3	poly(rC) binding protein 3 [Source:HGNC Symbol;Acc:HGNC:8651]	-0.77	0.00
151	ENSG00000161940	BCL6B	B-cell CLL/lymphoma 6, member B [Source:HGNC Symbol;Acc:HGNC:1002]	-0.76	0.00
152	ENSG00000065717	TLE2	transducin like enhancer of split 2 [Source:HGNC Symbol;Acc:HGNC:11838]	-0.76	0.00
153	ENSG00000173409	ARV1	ARV1 homolog, fatty acid homeostasis modulator [Source:HGNC Symbol;Acc:HGNC:29561]	-0.76	0.00
154	ENSG00000179546	HTR1D	5-hydroxytryptamine (serotonin) receptor 1D, G protein-coupled [Source:HGNC Symbol;Acc:HGNC:5289]	-0.76	0.00
155	ENSG00000214595	EML6	echinoderm microtubule associated protein like 6 [Source:HGNC Symbol;Acc:HGNC:35412]	-0.76	0.00
156	ENSG00000144134	RABL2A	RAB, member of RAS oncogene family-like 2A [Source:HGNC Symbol;Acc:HGNC:9799]	-0.76	0.00
157	ENSG00000001561	ENPP4	ectonucleotide pyrophosphatase/phosphodiesterase 4 (putative) [Source:HGNC Symbol;Acc:HGNC:3359]	-0.76	0.00
158	ENSG00000108439	PNPO	pyridoxamine 5'-phosphate oxidase [Source:HGNC Symbol;Acc:HGNC:30260]	-0.76	0.00
159	ENSG00000196391	ZNF774	zinc finger protein 774 [Source:HGNC Symbol;Acc:HGNC:33108]	-0.75	0.00
160	ENSG00000099958	DERL3	derlin 3 [Source:HGNC Symbol;Acc:HGNC:14236]	-0.75	0.00
161	ENSG00000116014	KISS1R	KISS1 receptor [Source:HGNC Symbol;Acc:HGNC:4510]	-0.75	0.00
162	ENSG00000253716	MINCR	MYC-induced long noncoding RNA [Source:HGNC Symbol;Acc:HGNC:51653]	-0.74	0.00
163	ENSG00000142875	PRKACB	protein kinase, cAMP-dependent, beta catalytic subunit [Source:HGNC Symbol;Acc:HGNC:9381]	-0.74	0.00
164	ENSG00000273604	C17orf96	chromosome 17 open reading frame 96 [Source:HGNC Symbol;Acc:HGNC:34493]	-0.74	0.01
165	ENSG00000116667	C1orf21	chromosome 1 open reading frame 21 [Source:HGNC Symbol;Acc:HGNC:15494]	-0.73	0.00
166	ENSG00000117477	CCDC181	coiled-coil domain containing 181 [Source:HGNC Symbol;Acc:HGNC:28051]	-0.73	0.00
167	ENSG00000142530	FAM71E1	family with sequence similarity 71 member E1 [Source:HGNC Symbol;Acc:HGNC:25107]	-0.73	0.00
168	ENSG00000004948	CALCR	calcitonin receptor [Source:HGNC Symbol;Acc:HGNC:1440]	-0.73	0.00
169	ENSG00000113522	RAD50	RAD50 homolog, double strand break repair protein [Source:HGNC Symbol;Acc:HGNC:9816]	-0.73	0.00
170	ENSG00000227640	SOX21-AS1	SOX21 antisense RNA 1 (head to head) [Source:HGNC Symbol;Acc:HGNC:39807]	-0.73	0.00
171	ENSG00000161544	CYGB	cytoglobin [Source:HGNC Symbol;Acc:HGNC:16505]	-0.73	0.00
172	ENSG00000161905	ALOX15	arachidonate 15-lipoxygenase [Source:HGNC Symbol;Acc:HGNC:433]	-0.72	0.01
173	ENSG00000103260	METRN	meteorin, glial cell differentiation regulator [Source:HGNC Symbol;Acc:HGNC:14151]	-0.72	0.00
174	ENSG00000204923	FBXO48	F-box protein 48 [Source:HGNC Symbol;Acc:HGNC:33857]	-0.72	0.00
175	ENSG00000189057	FAM111B	family with sequence similarity 111 member B [Source:HGNC Symbol;Acc:HGNC:24200]	-0.72	0.00
176	ENSG00000160716	CHRNB2	cholinergic receptor, nicotinic beta 2 [Source:HGNC Symbol;Acc:HGNC:1962]	-0.72	0.00

177	ENSG00000130489	SCO2	SCO2 cytochrome c oxidase assembly protein [Source:HGNC Symbol;Acc:HGNC:10604]	-0.71	0.00
178	ENSG00000018236	CNTN1	contactin 1 [Source:HGNC Symbol;Acc:HGNC:2171]	-0.71	0.00
179	ENSG00000124074	ENKD1	enkurin domain containing 1 [Source:HGNC Symbol;Acc:HGNC:25246]	-0.71	0.00
180	ENSG00000163888	CAMK2N2	calcium/calmodulin-dependent protein kinase II inhibitor 2 [Source:HGNC Symbol;Acc:HGNC:24197]	-0.70	0.00
181	ENSG00000197261	C6orf141	chromosome 6 open reading frame 141 [Source:HGNC Symbol;Acc:HGNC:21351]	-0.70	0.00
182	ENSG00000138111	MFSD13A	major facilitator superfamily domain containing 13A [Source:HGNC Symbol;Acc:HGNC:26196]	-0.70	0.00
183	ENSG00000104953	TLE6	transducin like enhancer of split 6 [Source:HGNC Symbol;Acc:HGNC:30788]	-0.70	0.00
184	ENSG00000198105	ZNF248	zinc finger protein 248 [Source:HGNC Symbol;Acc:HGNC:13041]	-0.69	0.00
185	ENSG00000163584	RPL22L1	ribosomal protein L22 like 1 [Source:HGNC Symbol;Acc:HGNC:27610]	-0.69	0.00
186	ENSG00000267640	CTD-2554C21.2		-0.69	0.00
187	ENSG00000167971	CASKIN1	CASK interacting protein 1 [Source:HGNC Symbol;Acc:HGNC:20879]	-0.69	0.00
188	ENSG00000197557	TTC30A	tetratricopeptide repeat domain 30A [Source:HGNC Symbol;Acc:HGNC:25853]	-0.69	0.00
189	ENSG00000204839	MROH6	maestro heat-like repeat family member 6 [Source:HGNC Symbol;Acc:HGNC:27814]	-0.69	0.00
190	ENSG00000188690	UROS	uroporphyrinogen III synthase [Source:HGNC Symbol;Acc:HGNC:12592]	-0.69	0.00
191	ENSG00000103034	NDRG4	NDRG family member 4 [Source:HGNC Symbol;Acc:HGNC:14466]	-0.68	0.00
192	ENSG00000114646	CSPG5	chondroitin sulfate proteoglycan 5 [Source:HGNC Symbol;Acc:HGNC:2467]	-0.68	0.00
193	ENSG00000113763	UNC5A	unc-5 netrin receptor A [Source:HGNC Symbol;Acc:HGNC:12567]	-0.68	0.00
194	ENSG00000112562	SMOC2	SPARC related modular calcium binding 2 [Source:HGNC Symbol;Acc:HGNC:20323]	-0.68	0.00
195	ENSG00000158106	RHPN1	rhophilin, Rho GTPase binding protein 1 [Source:HGNC Symbol;Acc:HGNC:19973]	-0.68	0.00
196	ENSG00000185697	MYBL1	v-myb avian myeloblastosis viral oncogene homolog-like 1 [Source:HGNC Symbol;Acc:HGNC:7547]	-0.68	0.00
197	ENSG00000110076	NRXN2	neurexin 2 [Source:HGNC Symbol;Acc:HGNC:8009]	-0.68	0.00
198	ENSG00000138769	CDKL2	cyclin-dependent kinase-like 2 (CDC2-related kinase) [Source:HGNC Symbol;Acc:HGNC:1782]	-0.68	0.00
199	ENSG00000141040	ZNF287	zinc finger protein 287 [Source:HGNC Symbol;Acc:HGNC:13502]	-0.67	0.00
200	ENSG00000166963	MAP1A	microtubule associated protein 1A [Source:HGNC Symbol;Acc:HGNC:6835]	-0.67	0.00
201	ENSG00000138380	CARF	calcium responsive transcription factor [Source:HGNC Symbol;Acc:HGNC:14435]	-0.67	0.00
202	ENSG00000102007	PLP2	proteolipid protein 2 (colonic epithelium-enriched) [Source:HGNC Symbol;Acc:HGNC:9087]	-0.67	0.00
203	ENSG00000174137	FAM53A	family with sequence similarity 53 member A [Source:HGNC Symbol;Acc:HGNC:31860]	-0.67	0.00
204	ENSG00000121486	TRMT1L	tRNA methyltransferase 1 like [Source:HGNC Symbol;Acc:HGNC:16782]	-0.67	0.00
205	ENSG00000180543	TSPYL5	TSPY-like 5 [Source:HGNC Symbol;Acc:HGNC:29367]	-0.67	0.00
206	ENSG00000132321	IQCA1	IQ motif containing with AAA domain 1 [Source:HGNC Symbol;Acc:HGNC:26195]	-0.67	0.00
207	ENSG00000196177	ACADSB	acyl-CoA dehydrogenase, short/branched chain [Source:HGNC Symbol;Acc:HGNC:91]	-0.67	0.00
208	ENSG00000101180	HRH3	histamine receptor H3 [Source:HGNC Symbol;Acc:HGNC:5184]	-0.67	0.01
209	ENSG00000168803	ADAL	adenosine deaminase-like [Source:HGNC Symbol;Acc:HGNC:31853]	-0.67	0.00
210	ENSG00000169169	CPT1C	carnitine palmitoyltransferase 1C [Source:HGNC Symbol;Acc:HGNC:18540]	-0.66	0.00
211	ENSG00000185818	NAT8L	N-acetyltransferase 8-like (GCN5-related, putative) [Source:HGNC Symbol;Acc:HGNC:26742]	-0.66	0.00
212	ENSG00000168056	LTBP3	latent transforming growth factor beta binding protein 3 [Source:HGNC Symbol;Acc:HGNC:6716]	-0.66	0.00

213	ENSG00000142961	MOB3C	MOB kinase activator 3C [Source:HGNC Symbol;Acc:HGNC:29800]	-0.66	0.00
214	ENSG00000135740	SLC9A5	solute carrier family 9, subfamily A (NHE5, cation proton antiporter 5), member 5 [Source:HGNC Symbol;Acc:HGNC:11078]	-0.66	0.00
215	ENSG00000115226	FNDC4	fibronectin type III domain containing 4 [Source:HGNC Symbol;Acc:HGNC:20239]	-0.66	0.00
216	ENSG00000189046	ALKBH2	alkB homolog 2, alpha-ketoglutarate-dependent dioxygenase [Source:HGNC Symbol;Acc:HGNC:32487]	-0.66	0.00
217	ENSG00000116254	CHD5	chromodomain helicase DNA binding protein 5 [Source:HGNC Symbol;Acc:HGNC:16816]	-0.66	0.00
218	ENSG00000008056	SYN1	synapsin I [Source:HGNC Symbol;Acc:HGNC:11494]	-0.66	0.00
219	ENSG00000140104	C14orf79	chromosome 14 open reading frame 79 [Source:HGNC Symbol;Acc:HGNC:20126]	-0.66	0.00
220	ENSG00000119703	ZC2HC1C	zinc finger, C2HC-type containing 1C [Source:HGNC Symbol;Acc:HGNC:20354]	-0.65	0.00
221	ENSG00000127585	FBXL16	F-box and leucine-rich repeat protein 16 [Source:HGNC Symbol;Acc:HGNC:14150]	-0.65	0.00
222	ENSG00000068831	RASGRP2	RAS guanyl releasing protein 2 (calcium and DAG-regulated) [Source:HGNC Symbol;Acc:HGNC:9879]	-0.65	0.00
223	ENSG00000100360	IFT27	intraflagellar transport 27 [Source:HGNC Symbol;Acc:HGNC:18626]	-0.65	0.00
224	ENSG00000149557	FEZ1	fasciculation and elongation protein zeta 1 [Source:HGNC Symbol;Acc:HGNC:3659]	-0.65	0.00
225	ENSG00000180616	SSTR2	somatostatin receptor 2 [Source:HGNC Symbol;Acc:HGNC:11331]	-0.65	0.00
226	ENSG00000166199	ALKBH3	alkB homolog 3, alpha-ketoglutarate-dependent dioxygenase [Source:HGNC Symbol;Acc:HGNC:30141]	-0.65	0.00
227	ENSG00000126882	FAM78A	family with sequence similarity 78 member A [Source:HGNC Symbol;Acc:HGNC:25465]	-0.65	0.00
228	ENSG00000139160	METTL20	methyltransferase like 20 [Source:HGNC Symbol;Acc:HGNC:28739]	-0.65	0.00
229	ENSG00000196544	BORCS6	BLOC-1 related complex subunit 6 [Source:HGNC Symbol;Acc:HGNC:25939]	-0.65	0.00
230	ENSG00000248746	ACTN3	actinin, alpha 3 (gene/pseudogene) [Source:HGNC Symbol;Acc:HGNC:165]	-0.65	0.00
231	ENSG00000280780	JAKMIP2-AS1	JAKMIP2 antisense RNA 1 [Source:HGNC Symbol;Acc:HGNC:27203]	-0.64	0.00
232	ENSG00000156973	PDE6D	phosphodiesterase 6D [Source:HGNC Symbol;Acc:HGNC:8788]	-0.64	0.00
233	ENSG00000068971	PPP2R5B	protein phosphatase 2 regulatory subunit B', beta [Source:HGNC Symbol;Acc:HGNC:9310]	-0.64	0.00
234	ENSG00000047634	SCML1	sex comb on midleg-like 1 (Drosophila) [Source:HGNC Symbol;Acc:HGNC:10580]	-0.64	0.00
235	ENSG00000105136	ZNF419	zinc finger protein 419 [Source:HGNC Symbol;Acc:HGNC:20648]	-0.64	0.00
236	ENSG00000100346	CACNA1I	calcium channel, voltage-dependent, T type, alpha 1I subunit [Source:HGNC Symbol;Acc:HGNC:1396]	-0.64	0.00
237	ENSG00000204248	COL11A2	collagen, type XI, alpha 2 [Source:HGNC Symbol;Acc:HGNC:2187]	-0.63	0.00
238	ENSG00000172733	PURG	purine-rich element binding protein G [Source:HGNC Symbol;Acc:HGNC:17930]	-0.63	0.00
239	ENSG00000131437	KIF3A	kinesin family member 3A [Source:HGNC Symbol;Acc:HGNC:6319]	-0.63	0.00
240	ENSG00000188760	TMEM198	transmembrane protein 198 [Source:HGNC Symbol;Acc:HGNC:33704]	-0.63	0.00
241	ENSG00000171130	ATP6V0E2	ATPase, H+ transporting V0 subunit e2 [Source:HGNC Symbol;Acc:HGNC:21723]	-0.63	0.00
242	ENSG00000115318	LOXL3	lysyl oxidase like 3 [Source:HGNC Symbol;Acc:HGNC:13869]	-0.63	0.00
243	ENSG00000180347	CCDC129	coiled-coil domain containing 129 [Source:HGNC Symbol;Acc:HGNC:27363]	-0.63	0.00
244	ENSG00000204642	HLA-F	major histocompatibility complex, class I, F [Source:HGNC Symbol;Acc:HGNC:4963]	-0.62	0.00
245	ENSG00000089723	OTUB2	OTU deubiquitinase, ubiquitin aldehyde binding 2 [Source:HGNC Symbol;Acc:HGNC:20351]	-0.62	0.00
246	ENSG00000198440	ZNF583	zinc finger protein 583 [Source:HGNC Symbol;Acc:HGNC:26427]	-0.62	0.00
247	ENSG00000059915	PSD	pleckstrin and Sec7 domain containing [Source:HGNC Symbol;Acc:HGNC:9507]	-0.62	0.00
248	ENSG00000103507	BCKDK	branched chain ketoacid dehydrogenase kinase [Source:HGNC Symbol;Acc:HGNC:16902]	-0.62	0.00

249	ENSG00000169598	DFFB	DNA fragmentation factor, 40kDa, beta polypeptide (caspase-activated DNase) [Source:HGNC Symbol;Acc:HGNC:2773]	-0.62	0.00
250	ENSG00000135439	AGAP2	ArfGAP with GTPase domain, ankyrin repeat and PH domain 2 [Source:HGNC Symbol;Acc:HGNC:16921]	-0.62	0.01
251	ENSG00000105270	CLIP3	CAP-Gly domain containing linker protein 3 [Source:HGNC Symbol;Acc:HGNC:24314]	-0.62	0.00
252	ENSG00000182405	PGBD4	piggyBac transposable element derived 4 [Source:HGNC Symbol;Acc:HGNC:19401]	-0.62	0.00
253	ENSG00000153982	GDPD1	glycerophosphodiester phosphodiesterase domain containing 1 [Source:HGNC Symbol;Acc:HGNC:20883]	-0.61	0.00
254	ENSG00000139364	TMEM132B	transmembrane protein 132B [Source:HGNC Symbol;Acc:HGNC:29397]	-0.61	0.00
255	ENSG00000117280	RAB29	RAB29, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:9789]	-0.61	0.00
256	ENSG00000256546	AC156455.1		-0.61	0.00
257	ENSG00000169026	MFSD7	major facilitator superfamily domain containing 7 [Source:HGNC Symbol;Acc:HGNC:26177]	-0.61	0.00
258	ENSG00000119547	ONECUT2	one cut homeobox 2 [Source:HGNC Symbol;Acc:HGNC:8139]	-0.61	0.00
259	ENSG00000249992	TMEM158	transmembrane protein 158 (gene/pseudogene) [Source:HGNC Symbol;Acc:HGNC:30293]	-0.61	0.00
260	ENSG00000142459	EVI5L	ecotropic viral integration site 5-like [Source:HGNC Symbol;Acc:HGNC:30464]	-0.60	0.00
261	ENSG00000148814	LRRC27	leucine rich repeat containing 27 [Source:HGNC Symbol;Acc:HGNC:29346]	-0.60	0.00
262	ENSG00000173673	HES3	hes family bHLH transcription factor 3 [Source:HGNC Symbol;Acc:HGNC:26226]	-0.60	0.00
263	ENSG00000273274	ZBTB8B	zinc finger and BTB domain containing 8B [Source:HGNC Symbol;Acc:HGNC:37057]	-0.59	0.00
264	ENSG00000205213	LGR4	leucine-rich repeat containing G protein-coupled receptor 4 [Source:HGNC Symbol;Acc:HGNC:13299]	-0.59	0.00
265	ENSG00000176022	B3GALT6	Beta-1,3-galactosyltransferase 6 [Source:HGNC Symbol;Acc:HGNC:17978]	-0.59	0.00
266	ENSG00000163945	UVSSA	UV stimulated scaffold protein A [Source:HGNC Symbol;Acc:HGNC:29304]	-0.59	0.00
267	ENSG00000141576	RNF157	ring finger protein 157 [Source:HGNC Symbol;Acc:HGNC:29402]	-0.59	0.00
268	ENSG00000100299	ARSA	arylsulfatase A [Source:HGNC Symbol;Acc:HGNC:713]	-0.59	0.00
269	ENSG00000120784	ZFP30	ZFP30 zinc finger protein [Source:HGNC Symbol;Acc:HGNC:29555]	-0.59	0.00
270	ENSG00000115255	REEP6	receptor accessory protein 6 [Source:HGNC Symbol;Acc:HGNC:30078]	-0.58	0.00
271	ENSG00000186300	ZNF555	zinc finger protein 555 [Source:HGNC Symbol;Acc:HGNC:28382]	-0.58	0.00
272	ENSG00000109618	SEPSECS	Sep (O-phosphoserine) tRNA:Sec (selenocysteine) tRNA synthase [Source:HGNC Symbol;Acc:HGNC:30605]	-0.58	0.00
273	ENSG00000166997	CNPY4	canopy FGF signaling regulator 4 [Source:HGNC Symbol;Acc:HGNC:28631]	-0.58	0.00
274	ENSG00000115257	PCSK4	proprotein convertase subtilisin/kexin type 4 [Source:HGNC Symbol;Acc:HGNC:8746]	-0.58	0.00
275	ENSG00000247516	MIR4458HG	MIR4458 host gene [Source:HGNC Symbol;Acc:HGNC:49008]	-0.58	0.00
276	ENSG00000100604	CHGA	chromogranin A [Source:HGNC Symbol;Acc:HGNC:1929]	-0.58	0.00
277	ENSG00000145439	CBR4	carbonyl reductase 4 [Source:HGNC Symbol;Acc:HGNC:25891]	-0.58	0.00
278	ENSG00000153093	ACOXL	acyl-CoA oxidase-like [Source:HGNC Symbol;Acc:HGNC:25621]	-0.58	0.00
279	ENSG00000151093	OXSM	3-oxoacyl-ACP synthase, mitochondrial [Source:HGNC Symbol;Acc:HGNC:26063]	-0.58	0.00
280	ENSG00000100078	PLA2G3	phospholipase A2 group III [Source:HGNC Symbol;Acc:HGNC:17934]	-0.57	0.00
281	ENSG00000137720	C11orf1	chromosome 11 open reading frame 1 [Source:HGNC Symbol;Acc:HGNC:1163]	-0.57	0.00
282	ENSG00000131370	SH3BP5	SH3-domain binding protein 5 (BTK-associated) [Source:HGNC Symbol;Acc:HGNC:10827]	-0.56	0.00
283	ENSG00000269404	SPIB	Spi-B transcription factor (Spi-1/PU.1 related) [Source:HGNC Symbol;Acc:HGNC:11242]	-0.56	0.00

284	ENSG00000091592	NLRP1	NLR family, pyrin domain containing 1 [Source:HGNC Symbol;Acc:HGNC:14374]	-0.56	0.00
285	ENSG00000163013	FBXO41	F-box protein 41 [Source:HGNC Symbol;Acc:HGNC:29409]	-0.56	0.00
286	ENSG00000006283	CACNA1G	calcium channel, voltage-dependent, T type, alpha 1G subunit [Source:HGNC Symbol;Acc:HGNC:1394]	-0.56	0.00
287	ENSG00000127419	TMEM175	transmembrane protein 175 [Source:HGNC Symbol;Acc:HGNC:28709]	-0.56	0.00
288	ENSG00000178053	MLF1	myeloid leukemia factor 1 [Source:HGNC Symbol;Acc:HGNC:7125]	-0.56	0.00
289	ENSG00000149929	HIRIP3	HIRA interacting protein 3 [Source:HGNC Symbol;Acc:HGNC:4917]	-0.56	0.00
290	ENSG00000187867	PALM3	paralemmin 3 [Source:HGNC Symbol;Acc:HGNC:33274]	-0.55	0.00
291	ENSG00000076258	FMO4	flavin containing monooxygenase 4 [Source:HGNC Symbol;Acc:HGNC:3772]	-0.55	0.00
292	ENSG00000205269	TMEM170B	transmembrane protein 170B [Source:HGNC Symbol;Acc:HGNC:34244]	-0.55	0.00
293	ENSG00000171456	ASXL1	additional sex combs like 1, transcriptional regulator [Source:HGNC Symbol;Acc:HGNC:18318]	-0.55	0.00
294	ENSG00000105357	MYH14	myosin, heavy chain 14, non-muscle [Source:HGNC Symbol;Acc:HGNC:23212]	-0.55	0.00
295	ENSG00000173442	EHBP1L1	EH domain binding protein 1 like 1 [Source:HGNC Symbol;Acc:HGNC:30682]	-0.55	0.00
296	ENSG00000005020	SKAP2	src kinase associated phosphoprotein 2 [Source:HGNC Symbol;Acc:HGNC:15687]	-0.55	0.00
297	ENSG00000018625	ATP1A2	ATPase, Na+/K+ transporting, alpha 2 polypeptide [Source:HGNC Symbol;Acc:HGNC:800]	-0.54	0.00
298	ENSG00000099949	LZTR1	leucine-zipper-like transcription regulator 1 [Source:HGNC Symbol;Acc:HGNC:6742]	-0.54	0.00
299	ENSG00000172086	KRCC1	lysine-rich coiled-coil 1 [Source:HGNC Symbol;Acc:HGNC:28039]	-0.54	0.00
300	ENSG00000167716	WDR81	WD repeat domain 81 [Source:HGNC Symbol;Acc:HGNC:26600]	-0.54	0.00
301	ENSG00000063854	HAGH	hydroxyacylglutathione hydrolase [Source:HGNC Symbol;Acc:HGNC:4805]	-0.54	0.00
302	ENSG00000073464	CLCN4	chloride channel, voltage-sensitive 4 [Source:HGNC Symbol;Acc:HGNC:2022]	-0.54	0.00
303	ENSG00000135617	PRADC1	protease-associated domain containing 1 [Source:HGNC Symbol;Acc:HGNC:16047]	-0.54	0.00
304	ENSG00000165215	CLDN3	claudin 3 [Source:HGNC Symbol;Acc:HGNC:2045]	-0.54	0.00
305	ENSG00000152133	GPATCH11	G-patch domain containing 11 [Source:HGNC Symbol;Acc:HGNC:26768]	-0.53	0.00
306	ENSG00000136052	SLC41A2	solute carrier family 41 (magnesium transporter), member 2 [Source:HGNC Symbol;Acc:HGNC:31045]	-0.53	0.01
307	ENSG00000130558	OLFM1	olfactomedin 1 [Source:HGNC Symbol;Acc:HGNC:17187]	-0.53	0.00
308	ENSG00000081377	CDC14B	cell division cycle 14B [Source:HGNC Symbol;Acc:HGNC:1719]	-0.53	0.00
309	ENSG00000246089	RP11-115C21.2		-0.53	0.00
310	ENSG00000103460	TOX3	TOX high mobility group box family member 3 [Source:HGNC Symbol;Acc:HGNC:11972]	-0.53	0.00
311	ENSG00000099625	CBARP	calcium channel, voltage-dependent, beta subunit associated regulatory protein [Source:HGNC Symbol;Acc:HGNC:28617]	-0.53	0.00
312	ENSG00000133460	SLC2A11	solute carrier family 2 (facilitated glucose transporter), member 11 [Source:HGNC Symbol;Acc:HGNC:14239]	-0.53	0.00
313	ENSG00000135473	PAN2	PAN2 poly(A) specific ribonuclease subunit [Source:HGNC Symbol;Acc:HGNC:20074]	-0.53	0.00
314	ENSG00000130707	ASS1	argininosuccinate synthase 1 [Source:HGNC Symbol;Acc:HGNC:758]	-0.52	0.00
315	ENSG00000164169	PRMT9	protein arginine methyltransferase 9 [Source:HGNC Symbol;Acc:HGNC:25099]	-0.52	0.00
316	ENSG00000164187	LMBRD2	LMBR1 domain containing 2 [Source:HGNC Symbol;Acc:HGNC:25287]	-0.52	0.00
317	ENSG00000138185	ENTPD1	ectonucleoside triphosphate diphosphohydrolase 1 [Source:HGNC Symbol;Acc:HGNC:3363]	-0.52	0.00
318	ENSG00000110328	GALNT18	polypeptide N-acetylgalactosaminyltransferase 18 [Source:HGNC Symbol;Acc:HGNC:30488]	-0.52	0.00

319	ENSG00000108384	RAD51C	RAD51 paralog C [Source:HGNC Symbol;Acc:HGNC:9820]	-0.52	0.00
320	ENSG00000215375	MYL5	myosin light chain 5 [Source:HGNC Symbol;Acc:HGNC:7586]	-0.51	0.00
321	ENSG00000120860	CCDC53	coiled-coil domain containing 53 [Source:HGNC Symbol;Acc:HGNC:24256]	-0.51	0.00
322	ENSG00000162576	MXRA8	matrix-remodelling associated 8 [Source:HGNC Symbol;Acc:HGNC:7542]	-0.51	0.00
323	ENSG00000241472	PTPRG-AS1	PTPRG antisense RNA 1 [Source:HGNC Symbol;Acc:HGNC:44638]	-0.51	0.00
324	ENSG00000099953	MMP11	matrix metallopeptidase 11 [Source:HGNC Symbol;Acc:HGNC:7157]	-0.51	0.01
325	ENSG00000183044	ABAT	4-aminobutyrate aminotransferase [Source:HGNC Symbol;Acc:HGNC:23]	-0.51	0.00
326	ENSG00000137133	HINT2	histidine triad nucleotide binding protein 2 [Source:HGNC Symbol;Acc:HGNC:18344]	-0.51	0.00
327	ENSG00000196372	ASB13	ankyrin repeat and SOCS box containing 13 [Source:HGNC Symbol;Acc:HGNC:19765]	-0.51	0.00
328	ENSG00000164252	AGGF1	angiogenic factor with G-patch and FHA domains 1 [Source:HGNC Symbol;Acc:HGNC:24684]	-0.51	0.00
329	ENSG00000008516	MMP25	matrix metallopeptidase 25 [Source:HGNC Symbol;Acc:HGNC:14246]	-0.51	0.00
330	ENSG00000025156	HSF2	heat shock transcription factor 2 [Source:HGNC Symbol;Acc:HGNC:5225]	-0.50	0.00
331	ENSG00000077150	NFKB2	nuclear factor of kappa light polypeptide gene enhancer in B-cells 2 (p49/p100) [Source:HGNC Symbol;Acc:HGNC:7795]	-0.50	0.00
332	ENSG00000169683	LRRC45	leucine rich repeat containing 45 [Source:HGNC Symbol;Acc:HGNC:28302]	-0.50	0.00
333	ENSG00000163626	COX18	COX18 cytochrome c oxidase assembly factor [Source:HGNC Symbol;Acc:HGNC:26801]	-0.50	0.00
334	ENSG00000188827	SLX4	SLX4 structure-specific endonuclease subunit [Source:HGNC Symbol;Acc:HGNC:23845]	-0.50	0.00
335	ENSG00000137941	TTLL7	tubulin tyrosine ligase like 7 [Source:HGNC Symbol;Acc:HGNC:26242]	-0.50	0.00
336	ENSG00000121057	AKAP1	A-kinase anchoring protein 1 [Source:HGNC Symbol;Acc:HGNC:367]	-0.50	0.00
337	ENSG00000120256	LRP11	LDL receptor related protein 11 [Source:HGNC Symbol;Acc:HGNC:16936]	-0.50	0.00
338	ENSG00000196263	ZNF471	zinc finger protein 471 [Source:HGNC Symbol;Acc:HGNC:23226]	-0.50	0.00
339	ENSG00000126259	KIRREL2	kin of IRRE like 2 (Drosophila) [Source:HGNC Symbol;Acc:HGNC:18816]	-0.50	0.01
340	ENSG00000149476	TKFC	triokinase/FMN cyclase [Source:HGNC Symbol;Acc:HGNC:24552]	-0.50	0.00
341	ENSG00000186104	CYP2R1	cytochrome P450 family 2 subfamily R member 1 [Source:HGNC Symbol;Acc:HGNC:20580]	-0.50	0.00
342	ENSG00000162755	KLHDC9	kelch domain containing 9 [Source:HGNC Symbol;Acc:HGNC:28489]	-0.50	0.00
343	ENSG00000156860	FBRS	fibrosin [Source:HGNC Symbol;Acc:HGNC:20442]	-0.49	0.00
344	ENSG00000107872	FBXL15	F-box and leucine-rich repeat protein 15 [Source:HGNC Symbol;Acc:HGNC:28155]	-0.49	0.00
345	ENSG00000178882	FAM101A	family with sequence similarity 101 member A [Source:HGNC Symbol;Acc:HGNC:27051]	-0.49	0.00
346	ENSG00000178882	FAM101A	family with sequence similarity 101 member A [Source:HGNC Symbol;Acc:HGNC:27051]	-0.49	0.00
347	ENSG00000107890	ANKRD26	ankyrin repeat domain 26 [Source:HGNC Symbol;Acc:HGNC:29186]	-0.49	0.00
348	ENSG00000092051	JPH4	junctophilin 4 [Source:HGNC Symbol;Acc:HGNC:20156]	-0.49	0.00
349	ENSG00000109084	TMEM97	transmembrane protein 97 [Source:HGNC Symbol;Acc:HGNC:28106]	-0.48	0.00
350	ENSG00000099910	KLHL22	kelch like family member 22 [Source:HGNC Symbol;Acc:HGNC:25888]	-0.48	0.00
351	ENSG00000140718	FTO	fat mass and obesity associated [Source:HGNC Symbol;Acc:HGNC:24678]	-0.48	0.00
352	ENSG00000168280	KIF5C	kinesin family member 5C [Source:HGNC Symbol;Acc:HGNC:6325]	-0.48	0.00
353	ENSG00000183837	PNMA3	paraneoplastic Ma antigen 3 [Source:HGNC Symbol;Acc:HGNC:18742]	-0.48	0.00
354	ENSG00000012822	CALCOCO1	calcium binding and coiled-coil domain 1 [Source:HGNC Symbol;Acc:HGNC:29306]	-0.48	0.00
355	ENSG00000205643	CDPF1	cysteine rich, DPF motif domain containing 1 [Source:HGNC Symbol;Acc:HGNC:33710]	-0.48	0.00

356	ENSG00000141736	ERBB2	erb-b2 receptor tyrosine kinase 2 [Source:HGNC Symbol;Acc:HGNC:3430]	-0.48	0.00
357	ENSG00000161267	BDH1	3-hydroxybutyrate dehydrogenase, type 1 [Source:HGNC Symbol;Acc:HGNC:1027]	-0.48	0.00
358	ENSG00000185129	PURA	purine-rich element binding protein A [Source:HGNC Symbol;Acc:HGNC:9701]	-0.48	0.00
359	ENSG00000164989	CCDC171	coiled-coil domain containing 171 [Source:HGNC Symbol;Acc:HGNC:29828]	-0.47	0.00
360	ENSG00000099365	STX1B	syntaxin 1B [Source:HGNC Symbol;Acc:HGNC:18539]	-0.47	0.00
361	ENSG00000167740	CYB5D2	cytochrome b5 domain containing 2 [Source:HGNC Symbol;Acc:HGNC:28471]	-0.47	0.00
362	ENSG00000141540	TTYH2	tweety family member 2 [Source:HGNC Symbol;Acc:HGNC:13877]	-0.47	0.00
363	ENSG00000010626	LRRC23	leucine rich repeat containing 23 [Source:HGNC Symbol;Acc:HGNC:19138]	-0.47	0.00
364	ENSG00000101210	EEF1A2	eukaryotic translation elongation factor 1 alpha 2 [Source:HGNC Symbol;Acc:HGNC:3192]	-0.47	0.00
365	ENSG00000006757	PNPLA4	patatin like phospholipase domain containing 4 [Source:HGNC Symbol;Acc:HGNC:24887]	-0.47	0.00
366	ENSG00000173264	GPR137	G protein-coupled receptor 137 [Source:HGNC Symbol;Acc:HGNC:24300]	-0.47	0.00
367	ENSG00000004455	AK2	adenylate kinase 2 [Source:HGNC Symbol;Acc:HGNC:362]	-0.47	0.00
368	ENSG00000176055	MBLAC2	metallo-beta-lactamase domain containing 2 [Source:HGNC Symbol;Acc:HGNC:33711]	-0.47	0.00
369	ENSG00000173705	SUSD5	sushi domain containing 5 [Source:HGNC Symbol;Acc:HGNC:29061]	-0.47	0.00
370	ENSG00000114473	IQCG	IQ motif containing G [Source:HGNC Symbol;Acc:HGNC:25251]	-0.47	0.00
371	ENSG00000128185	DGCR6L	DiGeorge syndrome critical region gene 6-like [Source:HGNC Symbol;Acc:HGNC:18551]	-0.46	0.00
372	ENSG00000018510	AGPS	alkylglycerone phosphate synthase [Source:HGNC Symbol;Acc:HGNC:327]	-0.46	0.00
373	ENSG00000166562	SEC11C	SEC11 homolog C, signal peptidase complex subunit [Source:HGNC Symbol;Acc:HGNC:23400]	-0.46	0.00
374	ENSG00000197763	TXNRD3	thioredoxin reductase 3 [Source:HGNC Symbol;Acc:HGNC:20667]	-0.46	0.00
375	ENSG00000197763	TXNRD3	thioredoxin reductase 3 [Source:HGNC Symbol;Acc:HGNC:20667]	-0.46	0.00
376	ENSG00000185483	ROR1	receptor tyrosine kinase-like orphan receptor 1 [Source:HGNC Symbol;Acc:HGNC:10256]	-0.46	0.00
377	ENSG00000072818	ACAP1	ArfGAP with coiled-coil, ankyrin repeat and PH domains 1 [Source:HGNC Symbol;Acc:HGNC:16467]	-0.46	0.00
378	ENSG00000204104	TRAF3IP1	TRAF3 interacting protein 1 [Source:HGNC Symbol;Acc:HGNC:17861]	-0.46	0.00
379	ENSG00000214655	ZSWIM8	zinc finger, SWIM-type containing 8 [Source:HGNC Symbol;Acc:HGNC:23528]	-0.46	0.00
380	ENSG00000204852	TCTN1	tectonic family member 1 [Source:HGNC Symbol;Acc:HGNC:26113]	-0.46	0.00
381	ENSG00000144749	LRIG1	leucine-rich repeats and immunoglobulin-like domains 1 [Source:HGNC Symbol;Acc:HGNC:17360]	-0.46	0.00
382	ENSG00000163344	PMVK	phosphomevalonate kinase [Source:HGNC Symbol;Acc:HGNC:9141]	-0.46	0.00
383	ENSG00000177239	MAN1B1	mannosidase, alpha, class 1B, member 1 [Source:HGNC Symbol;Acc:HGNC:6823]	-0.46	0.00
384	ENSG00000138604	GLCE	glucuronic acid epimerase [Source:HGNC Symbol;Acc:HGNC:17855]	-0.46	0.00
385	ENSG00000196507	TCEAL3	transcription elongation factor A (SII)-like 3 [Source:HGNC Symbol;Acc:HGNC:28247]	-0.45	0.00
386	ENSG00000166965	RCCD1	RCC1 domain containing 1 [Source:HGNC Symbol;Acc:HGNC:30457]	-0.45	0.00
387	ENSG00000022976	ZNF839	zinc finger protein 839 [Source:HGNC Symbol;Acc:HGNC:20345]	-0.45	0.00
388	ENSG00000175175	PPM1E	protein phosphatase, Mg2+/Mn2+ dependent 1E [Source:HGNC Symbol;Acc:HGNC:19322]	-0.45	0.00
389	ENSG00000168806	LCMT2	leucine carboxyl methyltransferase 2 [Source:HGNC Symbol;Acc:HGNC:17558]	-0.45	0.00
390	ENSG00000175854	SWI5	SWI5 homologous recombination repair protein [Source:HGNC Symbol;Acc:HGNC:31412]	-0.45	0.00

391	ENSG00000204682	CASC10	cancer susceptibility candidate 10 [Source:HGNC Symbol;Acc:HGNC:31448]	-0.45	0.00
392	ENSG00000167114	SLC27A4	solute carrier family 27 (fatty acid transporter), member 4 [Source:HGNC Symbol;Acc:HGNC:10998]	-0.45	0.00
393	ENSG00000167702	KIFC2	kinesin family member C2 [Source:HGNC Symbol;Acc:HGNC:29530]	-0.45	0.00
394	ENSG00000143578	CREB3L4	cAMP responsive element binding protein 3-like 4 [Source:HGNC Symbol;Acc:HGNC:18854]	-0.45	0.00
395	ENSG00000196456	ZNF775	zinc finger protein 775 [Source:HGNC Symbol;Acc:HGNC:28501]	-0.44	0.00
396	ENSG00000130635	COL5A1	collagen, type V, alpha 1 [Source:HGNC Symbol;Acc:HGNC:2209]	-0.44	0.00
397	ENSG00000137821	LRRC49	leucine rich repeat containing 49 [Source:HGNC Symbol;Acc:HGNC:25965]	-0.44	0.00
398	ENSG00000145029	NICN1	nicolin 1 [Source:HGNC Symbol;Acc:HGNC:18317]	-0.44	0.00
399	ENSG00000128228	SDF2L1	stromal cell-derived factor 2-like 1 [Source:HGNC Symbol;Acc:HGNC:10676]	-0.44	0.00
400	ENSG00000182325	FBXL6	F-box and leucine-rich repeat protein 6 [Source:HGNC Symbol;Acc:HGNC:13603]	-0.44	0.00
401	ENSG00000083123	BCKDHB	branched chain keto acid dehydrogenase E1, beta polypeptide [Source:HGNC Symbol;Acc:HGNC:987]	-0.44	0.00
402	ENSG00000184905	TCEAL2	transcription elongation factor A (SII)-like 2 [Source:HGNC Symbol;Acc:HGNC:29818]	-0.44	0.00
403	ENSG00000180921	FAM83H	family with sequence similarity 83 member H [Source:HGNC Symbol;Acc:HGNC:24797]	-0.44	0.00
404	ENSG00000196975	ANXA4	annexin A4 [Source:HGNC Symbol;Acc:HGNC:542]	-0.44	0.00
405	ENSG00000180902	D2HGDH	D-2-hydroxyglutarate dehydrogenase [Source:HGNC Symbol;Acc:HGNC:28358]	-0.44	0.00
406	ENSG00000131849	ZNF132	zinc finger protein 132 [Source:HGNC Symbol;Acc:HGNC:12916]	-0.43	0.00
407	ENSG00000187097	ENTPD5	ectonucleoside triphosphate diphosphohydrolase 5 [Source:HGNC Symbol;Acc:HGNC:3367]	-0.43	0.00
408	ENSG00000105255	FSD1	fibronectin type III and SPRY domain containing 1 [Source:HGNC Symbol;Acc:HGNC:13745]	-0.43	0.00
409	ENSG00000111879	FAM184A	family with sequence similarity 184 member A [Source:HGNC Symbol;Acc:HGNC:20991]	-0.43	0.00
410	ENSG00000169621	APLF	aprataxin and PNKP like factor [Source:HGNC Symbol;Acc:HGNC:28724]	-0.43	0.00
411	ENSG00000169621	APLF	aprataxin and PNKP like factor [Source:HGNC Symbol;Acc:HGNC:28724]	-0.43	0.00
412	ENSG00000160767	FAM189B	family with sequence similarity 189 member B [Source:HGNC Symbol;Acc:HGNC:1233]	-0.43	0.00
413	ENSG00000160216	AGPAT3	1-acylglycerol-3-phosphate O-acyltransferase 3 [Source:HGNC Symbol;Acc:HGNC:326]	-0.43	0.00
414	ENSG00000130751	NPAS1	neuronal PAS domain protein 1 [Source:HGNC Symbol;Acc:HGNC:7894]	-0.43	0.00
415	ENSG00000134901	KDELC1	KDEL motif containing 1 [Source:HGNC Symbol;Acc:HGNC:19350]	-0.43	0.00
416	ENSG00000139579	NABP2	nucleic acid binding protein 2 [Source:HGNC Symbol;Acc:HGNC:28412]	-0.43	0.00
417	ENSG00000204946	ZNF783	zinc finger family member 783 [Source:HGNC Symbol;Acc:HGNC:27222]	-0.43	0.00
418	ENSG00000054282	SDCCAG8	serologically defined colon cancer antigen 8 [Source:HGNC Symbol;Acc:HGNC:10671]	-0.43	0.00
419	ENSG00000146350	TBC1D32	TBC1 domain family member 32 [Source:HGNC Symbol;Acc:HGNC:21485]	-0.43	0.00
420	ENSG00000176058	TPRN	taperin [Source:HGNC Symbol;Acc:HGNC:26894]	-0.42	0.00
421	ENSG00000151338	MIPOL1	mirror-image polydactyly 1 [Source:HGNC Symbol;Acc:HGNC:21460]	-0.42	0.00
422	ENSG00000164114	MAP9	microtubule associated protein 9 [Source:HGNC Symbol;Acc:HGNC:26118]	-0.42	0.00
423	ENSG00000153094	BCL2L11	BCL2-like 11 (apoptosis facilitator) [Source:HGNC Symbol;Acc:HGNC:994]	-0.42	0.00
424	ENSG00000008300	CELSR3	cadherin, EGF LAG seven-pass G-type receptor 3 [Source:HGNC Symbol;Acc:HGNC:3230]	-0.42	0.00

425	ENSG00000147955	SIGMAR1	sigma non-opioid intracellular receptor 1 [Source:HGNC Symbol;Acc:HGNC:8157]	-0.42	0.00
426	ENSG00000235106	LINC00094	long intergenic non-protein coding RNA 94 [Source:HGNC Symbol;Acc:HGNC:24742]	-0.42	0.00
427	ENSG00000103150	MLYCD	malonyl-CoA decarboxylase [Source:HGNC Symbol;Acc:HGNC:7150]	-0.42	0.00
428	ENSG00000091844	RGS17	regulator of G-protein signaling 17 [Source:HGNC Symbol;Acc:HGNC:14088]	-0.42	0.00
429	ENSG00000140545	MFGE8	milk fat globule-EGF factor 8 protein [Source:HGNC Symbol;Acc:HGNC:7036]	-0.42	0.00
430	ENSG00000100592	DAAM1	dishevelled associated activator of morphogenesis 1 [Source:HGNC Symbol;Acc:HGNC:18142]	-0.42	0.00
431	ENSG00000134207	SYT6	synaptotagmin 6 [Source:HGNC Symbol;Acc:HGNC:18638]	-0.42	0.00
432	ENSG00000229544	NKX1-2	NK1 homeobox 2 [Source:HGNC Symbol;Acc:HGNC:31652]	-0.41	0.00
433	ENSG00000159921	GNE	glucosamine (UDP-N-acetyl)-2-epimerase/N-acetylmannosamine kinase [Source:HGNC Symbol;Acc:HGNC:23657]	-0.41	0.00
434	ENSG00000060642	PIGV	phosphatidylinositol glycan anchor biosynthesis class V [Source:HGNC Symbol;Acc:HGNC:26031]	-0.41	0.00
435	ENSG00000122512	PMS2	PMS1 homolog 2, mismatch repair system component [Source:HGNC Symbol;Acc:HGNC:9122]	-0.41	0.00
436	ENSG00000176473	WDR25	WD repeat domain 25 [Source:HGNC Symbol;Acc:HGNC:21064]	-0.41	0.00
437	ENSG00000151445	VIPAS39	VPS33B interacting protein, apical-basolateral polarity regulator, spe-39 homolog [Source:HGNC Symbol;Acc:HGNC:20347]	-0.41	0.00
438	ENSG00000172731	LRRC20	leucine rich repeat containing 20 [Source:HGNC Symbol;Acc:HGNC:23421]	-0.41	0.00
439	ENSG00000080709	KCNN2	potassium channel, calcium activated intermediate/small conductance subfamily N alpha, member 2 [Source:HGNC Symbol;Acc:HGNC:6291]	-0.41	0.00
440	ENSG00000164087	POC1A	POC1 centriolar protein A [Source:HGNC Symbol;Acc:HGNC:24488]	-0.41	0.00
441	ENSG00000179195	ZNF664	zinc finger protein 664 [Source:HGNC Symbol;Acc:HGNC:25406]	-0.41	0.00
442	ENSG00000182093	WRB	tryptophan rich basic protein [Source:HGNC Symbol;Acc:HGNC:12790]	-0.41	0.00
443	ENSG00000157911	PEX10	peroxisomal biogenesis factor 10 [Source:HGNC Symbol;Acc:HGNC:8851]	-0.40	0.00
444	ENSG00000095321	CRAT	carnitine O-acetyltransferase [Source:HGNC Symbol;Acc:HGNC:2342]	-0.40	0.00
445	ENSG00000167925	GHDC	GH3 domain containing [Source:HGNC Symbol;Acc:HGNC:24438]	-0.40	0.00
446	ENSG00000167654	ATCAY	ataxia, cerebellar, Cayman type [Source:HGNC Symbol;Acc:HGNC:779]	-0.40	0.00
447	ENSG00000116198	CEP104	centrosomal protein 104kDa [Source:HGNC Symbol;Acc:HGNC:24866]	-0.40	0.00
448	ENSG00000144792	ZNF660	zinc finger protein 660 [Source:HGNC Symbol;Acc:HGNC:26720]	-0.40	0.00
449	ENSG00000158019	BRE	brain and reproductive organ-expressed (TNFRSF1A modulator) [Source:HGNC Symbol;Acc:HGNC:1106]	-0.40	0.00
450	ENSG00000101166	PRELID3B	PRELI domain containing 3B [Source:HGNC Symbol;Acc:HGNC:15892]	-0.40	0.00
451	ENSG00000166105	GLB1L3	galactosidase beta 1 like 3 [Source:HGNC Symbol;Acc:HGNC:25147]	-0.40	0.00
452	ENSG00000025039	RRAGD	Ras-related GTP binding D [Source:HGNC Symbol;Acc:HGNC:19903]	-0.40	0.00
453	ENSG00000129473	BCL2L2	BCL2-like 2 [Source:HGNC Symbol;Acc:HGNC:995]	-0.40	0.00
454	ENSG00000185437	SH3BGR	SH3 domain binding glutamate-rich protein [Source:HGNC Symbol;Acc:HGNC:10822]	-0.40	0.00
455	ENSG00000162065	TBC1D24	TBC1 domain family member 24 [Source:HGNC Symbol;Acc:HGNC:29203]	-0.40	0.00
456	ENSG00000163867	ZMYM6	zinc finger, MYM-type 6 [Source:HGNC Symbol;Acc:HGNC:13050]	-0.40	0.00
457	ENSG00000171608	PIK3CD	phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit delta [Source:HGNC Symbol;Acc:HGNC:8977]	-0.40	0.00
458	ENSG00000131697	NPHP4	nephronophthisis 4 [Source:HGNC Symbol;Acc:HGNC:19104]	-0.40	0.00
459	ENSG00000171766	GATM	glycine amidinotransferase (L-arginine:glycine amidinotransferase) [Source:HGNC Symbol;Acc:HGNC:4175]	-0.40	0.00

460	ENSG00000153012	LGI2	leucine-rich repeat LGI family member 2 [Source:HGNC Symbol;Acc:HGNC:18710]	-0.40	0.00
461	ENSG00000050426	LETMD1	LETM1 domain containing 1 [Source:HGNC Symbol;Acc:HGNC:24241]	-0.40	0.00
462	ENSG00000164402	SEPT8	septin 8 [Source:HGNC Symbol;Acc:HGNC:16511]	-0.39	0.00
463	ENSG00000124374	PAIP2B	poly(A) binding protein interacting protein 2B [Source:HGNC Symbol;Acc:HGNC:29200]	-0.39	0.00
464	ENSG00000166750	SLFN5	schlafen family member 5 [Source:HGNC Symbol;Acc:HGNC:28286]	-0.39	0.00
465	ENSG00000144647	POMGNT2	protein O-linked mannose N-acetylglucosaminyltransferase 2 (beta 1,4-) [Source:HGNC Symbol;Acc:HGNC:25902]	-0.39	0.00
466	ENSG00000118855	MFSD1	major facilitator superfamily domain containing 1 [Source:HGNC Symbol;Acc:HGNC:25874]	-0.39	0.00
467	ENSG00000169750	RAC3	ras-related C3 botulinum toxin substrate 3 (rho family, small GTP binding protein Rac3) [Source:HGNC Symbol;Acc:HGNC:9803]	-0.39	0.00
468	ENSG00000134330	IAH1	isoamyl acetate-hydrolyzing esterase 1 homolog [Source:HGNC Symbol;Acc:HGNC:27696]	-0.39	0.00
469	ENSG00000144426	NBEAL1	neurobeachin like 1 [Source:HGNC Symbol;Acc:HGNC:20681]	-0.39	0.00
470	ENSG00000148408	CACNA1B	calcium channel, voltage-dependent, N type, alpha 1B subunit [Source:HGNC Symbol;Acc:HGNC:1389]	-0.38	0.00
471	ENSG00000164105	SAP30	Sin3A associated protein 30kDa [Source:HGNC Symbol;Acc:HGNC:10532]	-0.38	0.00
472	ENSG00000214941	ZSWIM7	zinc finger, SWIM-type containing 7 [Source:HGNC Symbol;Acc:HGNC:26993]	-0.38	0.00
473	ENSG00000137216	TMEM63B	transmembrane protein 63B [Source:HGNC Symbol;Acc:HGNC:17735]	-0.38	0.00
474	ENSG00000160326	SLC2A6	solute carrier family 2 (facilitated glucose transporter), member 6 [Source:HGNC Symbol;Acc:HGNC:11011]	-0.38	0.00
475	ENSG00000172638	EFEMP2	EGF containing fibulin-like extracellular matrix protein 2 [Source:HGNC Symbol;Acc:HGNC:3219]	-0.38	0.00
476	ENSG00000198707	CEP290	centrosomal protein 290kDa [Source:HGNC Symbol;Acc:HGNC:29021]	-0.38	0.00
477	ENSG00000131584	ACAP3	ArfGAP with coiled-coil, ankyrin repeat and PH domains 3 [Source:HGNC Symbol;Acc:HGNC:16754]	-0.38	0.00
478	ENSG00000079432	CIC	capicua transcriptional repressor [Source:HGNC Symbol;Acc:HGNC:14214]	-0.38	0.00
479	ENSG00000155016	CYP2U1	cytochrome P450 family 2 subfamily U member 1 [Source:HGNC Symbol;Acc:HGNC:20582]	-0.38	0.00
480	ENSG00000188130	MAPK12	mitogen-activated protein kinase 12 [Source:HGNC Symbol;Acc:HGNC:6874]	-0.38	0.00
481	ENSG00000099940	SNAP29	synaptosome associated protein 29kDa [Source:HGNC Symbol;Acc:HGNC:11133]	-0.38	0.00
482	ENSG00000008869	HEATR5B	HEAT repeat containing 5B [Source:HGNC Symbol;Acc:HGNC:29273]	-0.38	0.00
483	ENSG00000196700	ZNF512B	zinc finger protein 512B [Source:HGNC Symbol;Acc:HGNC:29212]	-0.38	0.00
484	ENSG00000161091	MFSD12	major facilitator superfamily domain containing 12 [Source:HGNC Symbol;Acc:HGNC:28299]	-0.37	0.00
485	ENSG00000163001	CFAP36	cilia and flagella associated protein 36 [Source:HGNC Symbol;Acc:HGNC:30540]	-0.37	0.00
486	ENSG00000133138	TBC1D8B	TBC1 domain family member 8B [Source:HGNC Symbol;Acc:HGNC:24715]	-0.37	0.00
487	ENSG00000071859	FAM50A	family with sequence similarity 50 member A [Source:HGNC Symbol;Acc:HGNC:18786]	-0.37	0.00
488	ENSG00000196922	ZNF252P	zinc finger protein 252, pseudogene [Source:HGNC Symbol;Acc:HGNC:13046]	-0.37	0.00
489	ENSG00000081177	EXD2	exonuclease 3'-5' domain containing 2 [Source:HGNC Symbol;Acc:HGNC:20217]	-0.37	0.00
490	ENSG00000014138	POLA2	polymerase (DNA directed), alpha 2, accessory subunit [Source:HGNC Symbol;Acc:HGNC:30073]	-0.37	0.00
491	ENSG00000108465	CDK5RAP3	CDK5 regulatory subunit associated protein 3 [Source:HGNC Symbol;Acc:HGNC:18673]	-0.37	0.00

492	ENSG00000147813	NAPRT	nicotinate phosphoribosyltransferase [Source:HGNC Symbol;Acc:HGNC:30450]	-0.37	0.00
493	ENSG00000100116	GCAT	glycine C-acetyltransferase [Source:HGNC Symbol;Acc:HGNC:4188]	-0.37	0.00
494	ENSG00000159423	ALDH4A1	aldehyde dehydrogenase 4 family member A1 [Source:HGNC Symbol;Acc:HGNC:406]	-0.37	0.00
495	ENSG00000047188	YTHDC2	YTH domain containing 2 [Source:HGNC Symbol;Acc:HGNC:24721]	-0.36	0.00
496	ENSG00000166347	CYB5A	cytochrome b5 type A (microsomal) [Source:HGNC Symbol;Acc:HGNC:2570]	-0.36	0.00
497	ENSG00000253738	OTUD6B-AS1	OTUD6B antisense RNA 1 (head to head) [Source:HGNC Symbol;Acc:HGNC:50466]	-0.36	0.00
498	ENSG00000184207	PGP	phosphoglycolate phosphatase [Source:HGNC Symbol;Acc:HGNC:8909]	-0.36	0.00
499	ENSG00000141971	MVB12A	multivesicular body subunit 12A [Source:HGNC Symbol;Acc:HGNC:25153]	-0.36	0.00
500	ENSG00000114388	NPRL2	NPR2-like, GATOR1 complex subunit [Source:HGNC Symbol;Acc:HGNC:24969]	-0.36	0.00
501	ENSG00000159363	ATP13A2	ATPase type 13A2 [Source:HGNC Symbol;Acc:HGNC:30213]	-0.36	0.00
502	ENSG00000074211	PPP2R2C	protein phosphatase 2 regulatory subunit B, gamma [Source:HGNC Symbol;Acc:HGNC:9306]	-0.36	0.00
503	ENSG00000254685	FPGT	fucose-1-phosphate guanylyltransferase [Source:HGNC Symbol;Acc:HGNC:3825]	-0.36	0.00
504	ENSG00000146834	MEPCE	methylphosphate capping enzyme [Source:HGNC Symbol;Acc:HGNC:20247]	-0.36	0.00
505	ENSG00000168243	GNG4	guanine nucleotide binding protein (G protein), gamma 4 [Source:HGNC Symbol;Acc:HGNC:4407]	-0.36	0.00
506	ENSG00000172915	NBEA	neurobeachin [Source:HGNC Symbol;Acc:HGNC:7648]	-0.36	0.00
507	ENSG00000151500	THYN1	thymocyte nuclear protein 1 [Source:HGNC Symbol;Acc:HGNC:29560]	-0.36	0.00
508	ENSG00000132406	TMEM128	transmembrane protein 128 [Source:HGNC Symbol;Acc:HGNC:28201]	-0.36	0.00
509	ENSG00000234787	LINC00458	long intergenic non-protein coding RNA 458 [Source:HGNC Symbol;Acc:HGNC:42807]	-0.36	0.00
510	ENSG00000155850	SLC26A2	solute carrier family 26 (anion exchanger), member 2 [Source:HGNC Symbol;Acc:HGNC:10994]	-0.35	0.00
511	ENSG00000101220	C20orf27	chromosome 20 open reading frame 27 [Source:HGNC Symbol;Acc:HGNC:15873]	-0.35	0.00
512	ENSG00000160284	SPATC1L	spermatogenesis and centriole associated 1-like [Source:HGNC Symbol;Acc:HGNC:1298]	-0.35	0.00
513	ENSG00000141759	TXNL4A	thioredoxin like 4A [Source:HGNC Symbol;Acc:HGNC:30551]	-0.35	0.00
514	ENSG00000033627	ATP6V0A1	ATPase, H+ transporting, lysosomal V0 subunit a1 [Source:HGNC Symbol;Acc:HGNC:865]	-0.35	0.00
515	ENSG00000117408	IPO13	importin 13 [Source:HGNC Symbol;Acc:HGNC:16853]	-0.35	0.00
516	ENSG00000143224	PPOX	protoporphyrinogen oxidase [Source:HGNC Symbol;Acc:HGNC:9280]	-0.35	0.00
517	ENSG00000131844	MCCC2	methylcrotonoyl-CoA carboxylase 2 [Source:HGNC Symbol;Acc:HGNC:6937]	-0.35	0.00
518	ENSG00000155876	RRAGA	Ras-related GTP binding A [Source:HGNC Symbol;Acc:HGNC:16963]	-0.35	0.00
519	ENSG00000064687	ABCA7	ATP binding cassette subfamily A member 7 [Source:HGNC Symbol;Acc:HGNC:37]	-0.35	0.00
520	ENSG00000169189	NSMCE1	NSE1 homolog, SMC5-SMC6 complex component [Source:HGNC Symbol;Acc:HGNC:29897]	-0.35	0.00
521	ENSG00000090971	NAT14	N-acetyltransferase 14 (GCN5-related, putative) [Source:HGNC Symbol;Acc:HGNC:28918]	-0.35	0.00
522	ENSG00000150867	PIP4K2A	phosphatidylinositol-5-phosphate 4-kinase, type II, alpha [Source:HGNC Symbol;Acc:HGNC:8997]	-0.35	0.00
523	ENSG00000188811	NHLRC3	NHL repeat containing 3 [Source:HGNC Symbol;Acc:HGNC:33751]	-0.35	0.00
524	ENSG00000153406	NMRAL1	NmrA-like family domain containing 1 [Source:HGNC Symbol;Acc:HGNC:24987]	-0.35	0.00
525	ENSG00000103494	RPGRIP1L	RPGRIP1-like [Source:HGNC Symbol;Acc:HGNC:29168]	-0.35	0.00
526	ENSG00000067836	ROGDI	rogdi homolog [Source:HGNC Symbol;Acc:HGNC:29478]	-0.35	0.00
527	ENSG00000091542	ALKBH5	alkB homolog 5, RNA demethylase [Source:HGNC Symbol;Acc:HGNC:25996]	-0.35	0.00

528	ENSG00000135924	DNAJB2	DnaJ heat shock protein family (Hsp40) member B2 [Source:HGNC Symbol;Acc:HGNC:5228]	-0.35	0.00
529	ENSG00000117758	STX12	syntaxin 12 [Source:HGNC Symbol;Acc:HGNC:11430]	-0.35	0.00
530	ENSG00000143314	MRPL24	mitochondrial ribosomal protein L24 [Source:HGNC Symbol;Acc:HGNC:14037]	-0.34	0.00
531	ENSG00000255302	EID1	EP300 interacting inhibitor of differentiation 1 [Source:HGNC Symbol;Acc:HGNC:1191]	-0.34	0.00
532	ENSG00000186416	NKRF	NFKB repressing factor [Source:HGNC Symbol;Acc:HGNC:19374]	-0.34	0.00
533	ENSG00000088888	MAVS	mitochondrial antiviral signaling protein [Source:HGNC Symbol;Acc:HGNC:29233]	-0.34	0.00
534	ENSG00000159079	C21orf59	chromosome 21 open reading frame 59 [Source:HGNC Symbol;Acc:HGNC:1301]	-0.34	0.00
535	ENSG00000159753	RLTPR	RGD motif, leucine rich repeats, tropomodulin domain and proline-rich containing [Source:HGNC Symbol;Acc:HGNC:27089]	-0.34	0.00
536	ENSG00000162066	AMDHD2	amidohydrolase domain containing 2 [Source:HGNC Symbol;Acc:HGNC:24262]	-0.34	0.00
537	ENSG00000111652	COPS7A	COP9 signalosome subunit 7A [Source:HGNC Symbol;Acc:HGNC:16758]	-0.34	0.00
538	ENSG00000142528	ZNF473	zinc finger protein 473 [Source:HGNC Symbol;Acc:HGNC:23239]	-0.34	0.00
539	ENSG00000014824	SLC30A9	solute carrier family 30 (zinc transporter), member 9 [Source:HGNC Symbol;Acc:HGNC:1329]	-0.34	0.00
540	ENSG00000103148	NPRL3	NPR3-like, GATOR1 complex subunit [Source:HGNC Symbol;Acc:HGNC:14124]	-0.34	0.00
541	ENSG00000149809	TM7SF2	transmembrane 7 superfamily member 2 [Source:HGNC Symbol;Acc:HGNC:11863]	-0.34	0.00
542	ENSG00000107874	CUEDC2	CUE domain containing 2 [Source:HGNC Symbol;Acc:HGNC:28352]	-0.34	0.00
543	ENSG00000149499	EML3	echinoderm microtubule associated protein like 3 [Source:HGNC Symbol;Acc:HGNC:26666]	-0.33	0.00
544	ENSG00000101443	WFDC2	WAP four-disulfide core domain 2 [Source:HGNC Symbol;Acc:HGNC:15939]	-0.33	0.00
545	ENSG00000169738	DCXR	dicarbonyl/L-xylulose reductase [Source:HGNC Symbol;Acc:HGNC:18985]	-0.33	0.00
546	ENSG00000101350	KIF3B	kinesin family member 3B [Source:HGNC Symbol;Acc:HGNC:6320]	-0.33	0.00
547	ENSG00000183010	PYCR1	pyrroline-5-carboxylate reductase 1 [Source:HGNC Symbol;Acc:HGNC:9721]	-0.33	0.01
548	ENSG00000198690	FAN1	FANCD2/FANCI-associated nuclease 1 [Source:HGNC Symbol;Acc:HGNC:29170]	-0.33	0.00
549	ENSG00000100418	DESI1	desumoylating isopeptidase 1 [Source:HGNC Symbol;Acc:HGNC:24577]	-0.33	0.00
550	ENSG00000169992	NLGN2	neuroligin 2 [Source:HGNC Symbol;Acc:HGNC:14290]	-0.33	0.00
551	ENSG00000116406	EDEM3	ER degradation enhancer, mannosidase alpha-like 3 [Source:HGNC Symbol;Acc:HGNC:16787]	-0.33	0.00
552	ENSG00000189060	H1F0	H1 histone family member 0 [Source:HGNC Symbol;Acc:HGNC:4714]	-0.33	0.00
553	ENSG00000128524	ATP6V1F	ATPase, H+ transporting, lysosomal 14kDa, V1 subunit F [Source:HGNC Symbol;Acc:HGNC:16832]	-0.33	0.00
554	ENSG00000136436	CALCOCO2	calcium binding and coiled-coil domain 2 [Source:HGNC Symbol;Acc:HGNC:29912]	-0.33	0.00
555	ENSG00000170876	TMEM43	transmembrane protein 43 [Source:HGNC Symbol;Acc:HGNC:28472]	-0.33	0.00
556	ENSG00000182963	GJC1	gap junction protein gamma 1 [Source:HGNC Symbol;Acc:HGNC:4280]	-0.33	0.00
557	ENSG00000009780	FAM76A	family with sequence similarity 76 member A [Source:HGNC Symbol;Acc:HGNC:28530]	-0.33	0.00
558	ENSG00000133612	AGAP3	ArfGAP with GTPase domain, ankyrin repeat and PH domain 3 [Source:HGNC Symbol;Acc:HGNC:16923]	-0.33	0.00
559	ENSG00000158079	PTPDC1	protein tyrosine phosphatase domain containing 1 [Source:HGNC Symbol;Acc:HGNC:30184]	-0.32	0.00
560	ENSG00000205060	SLC35B4	solute carrier family 35 (UDP-xylose/UDP-N-acetylglucosamine transporter), member B4 [Source:HGNC Symbol;Acc:HGNC:20584]	-0.32	0.00
561	ENSG00000069998	CECR5	cat eye syndrome chromosome region, candidate 5 [Source:HGNC Symbol;Acc:HGNC:1843]	-0.32	0.00

562	ENSG00000129317	PUS7L	pseudouridylate synthase 7-like [Source:HGNC Symbol;Acc:HGNC:25276]	-0.32	0.00
563	ENSG00000153130	SCOC	short coiled-coil protein [Source:HGNC Symbol;Acc:HGNC:20335]	-0.32	0.00
564	ENSG00000140105	WARS	tryptophanyl-tRNA synthetase [Source:HGNC Symbol;Acc:HGNC:12729]	-0.32	0.00
565	ENSG00000227372	TP73-AS1	TP73 antisense RNA 1 [Source:HGNC Symbol;Acc:HGNC:29052]	-0.32	0.00
566	ENSG00000135317	SNX14	sorting nexin 14 [Source:HGNC Symbol;Acc:HGNC:14977]	-0.32	0.00
567	ENSG00000137601	NEK1	NIMA-related kinase 1 [Source:HGNC Symbol;Acc:HGNC:7744]	-0.32	0.00
568	ENSG00000094916	CBX5	chromobox 5 [Source:HGNC Symbol;Acc:HGNC:1555]	-0.32	0.00
569	ENSG00000198056	PRIM1	primase, DNA, polypeptide 1 (49kDa) [Source:HGNC Symbol;Acc:HGNC:9369]	-0.31	0.00
570	ENSG00000110060	PUS3	pseudouridylate synthase 3 [Source:HGNC Symbol;Acc:HGNC:25461]	-0.31	0.00
571	ENSG00000188428	BLOC1S5	biogenesis of lysosomal organelles complex-1, subunit 5, muted [Source:HGNC Symbol;Acc:HGNC:18561]	-0.31	0.00
572	ENSG00000064545	TMEM161A	transmembrane protein 161A [Source:HGNC Symbol;Acc:HGNC:26020]	-0.31	0.00
573	ENSG00000196976	LAGE3	L antigen family member 3 [Source:HGNC Symbol;Acc:HGNC:26058]	-0.31	0.00
574	ENSG00000132024	CC2D1A	coiled-coil and C2 domain containing 1A [Source:HGNC Symbol;Acc:HGNC:30237]	-0.30	0.00
575	ENSG00000126391	FRMD8	FERM domain containing 8 [Source:HGNC Symbol;Acc:HGNC:25462]	-0.30	0.00
576	ENSG00000153767	GTF2E1	general transcription factor IIE subunit 1 [Source:HGNC Symbol;Acc:HGNC:4650]	-0.30	0.00
577	ENSG00000116984	MTR	5-methyltetrahydrofolate-homocysteine methyltransferase [Source:HGNC Symbol;Acc:HGNC:7468]	-0.30	0.00
578	ENSG00000162600	OMA1	OMA1 zinc metallopeptidase [Source:HGNC Symbol;Acc:HGNC:29661]	-0.30	0.00
579	ENSG00000170100	ZNF778	zinc finger protein 778 [Source:HGNC Symbol;Acc:HGNC:26479]	-0.30	0.00
580	ENSG00000185115	NSMCE3	NSE3 homolog, SMC5-SMC6 complex component [Source:HGNC Symbol;Acc:HGNC:7677]	-0.30	0.00
581	ENSG00000068885	IFT80	intraflagellar transport 80 [Source:HGNC Symbol;Acc:HGNC:29262]	-0.29	0.00
582	ENSG00000082146	STRADB	STE20-related kinase adaptor beta [Source:HGNC Symbol;Acc:HGNC:13205]	-0.29	0.00
583	ENSG00000151498	ACAD8	acyl-CoA dehydrogenase family member 8 [Source:HGNC Symbol;Acc:HGNC:87]	-0.29	0.00
584	ENSG00000197879	MYO1C	myosin IC [Source:HGNC Symbol;Acc:HGNC:7597]	-0.29	0.00
585	ENSG00000163832	ELP6	elongator acetyltransferase complex subunit 6 [Source:HGNC Symbol;Acc:HGNC:25976]	-0.29	0.00
586	ENSG00000154124	OTULIN	OTU deubiquitinase with linear linkage specificity [Source:HGNC Symbol;Acc:HGNC:25118]	-0.29	0.00
587	ENSG00000129003	VPS13C	vacuolar protein sorting 13 homolog C (S. cerevisiae) [Source:HGNC Symbol;Acc:HGNC:23594]	-0.29	0.00
588	ENSG00000196437	ZNF569	zinc finger protein 569 [Source:HGNC Symbol;Acc:HGNC:24737]	-0.28	0.00
589	ENSG00000147162	OGT	O-linked N-acetylglucosamine (GlcNAc) transferase [Source:HGNC Symbol;Acc:HGNC:8127]	-0.28	0.00
590	ENSG00000158457	TSPAN33	tetraspanin 33 [Source:HGNC Symbol;Acc:HGNC:28743]	-0.28	0.00
591	ENSG00000128294	TPST2	tyrosylprotein sulfotransferase 2 [Source:HGNC Symbol;Acc:HGNC:12021]	-0.28	0.00
592	ENSG00000105419	MEIS3	Meis homeobox 3 [Source:HGNC Symbol;Acc:HGNC:29537]	-0.28	0.00
593	ENSG00000056097	ZFR	zinc finger RNA binding protein [Source:HGNC Symbol;Acc:HGNC:17277]	0.30	0.00
594	ENSG00000164611	PTTG1	pituitary tumor-transforming 1 [Source:HGNC Symbol;Acc:HGNC:9690]	0.30	0.00
595	ENSG00000214113	LYRM4	LYR motif containing 4 [Source:HGNC Symbol;Acc:HGNC:21365]	0.30	0.00
596	ENSG00000168615	ADAM9	ADAM metallopeptidase domain 9 [Source:HGNC Symbol;Acc:HGNC:216]	0.31	0.00
597	ENSG00000148672	GLUD1	glutamate dehydrogenase 1 [Source:HGNC Symbol;Acc:HGNC:4335]	0.31	0.00
598	ENSG00000060339	CCAR1	cell division cycle and apoptosis regulator 1 [Source:HGNC Symbol;Acc:HGNC:24236]	0.31	0.00
599	ENSG00000100625	SIX4	SIX homeobox 4 [Source:HGNC Symbol;Acc:HGNC:10890]	0.31	0.00

600	ENSG00000092068	SLC7A8	solute carrier family 7 (amino acid transporter light chain, L system), member 8 [Source:HGNC Symbol;Acc:HGNC:11066]	0.31	0.00
601	ENSG00000175061	LRRC75A-AS1	LRRC75A antisense RNA 1 [Source:HGNC Symbol;Acc:HGNC:28619]	0.31	0.00
602	ENSG00000269893	SNHG8	small nucleolar RNA host gene 8 [Source:HGNC Symbol;Acc:HGNC:33098]	0.31	0.00
603	ENSG00000134755	DSC2	desmocollin 2 [Source:HGNC Symbol;Acc:HGNC:3036]	0.31	0.00
604	ENSG00000112306	RPS12	ribosomal protein S12 [Source:HGNC Symbol;Acc:HGNC:10385]	0.32	0.00
605	ENSG00000001036	FUCA2	fucosidase, alpha-L- 2, plasma [Source:HGNC Symbol;Acc:HGNC:4008]	0.32	0.00
606	ENSG00000150471	ADGRL3	adhesion G protein-coupled receptor L3 [Source:HGNC Symbol;Acc:HGNC:20974]	0.32	0.00
607	ENSG00000113140	SPARC	secreted protein, acidic, cysteine-rich (osteonectin) [Source:HGNC Symbol;Acc:HGNC:11219]	0.32	0.00
608	ENSG00000064115	TM7SF3	transmembrane 7 superfamily member 3 [Source:HGNC Symbol;Acc:HGNC:23049]	0.32	0.00
609	ENSG00000139112	GABARAPL1	GABA(A) receptor-associated protein like 1 [Source:HGNC Symbol;Acc:HGNC:4068]	0.32	0.00
610	ENSG00000186767	SPIN4	spindlin family member 4 [Source:HGNC Symbol;Acc:HGNC:27040]	0.33	0.00
611	ENSG00000167671	UBXN6	UBX domain protein 6 [Source:HGNC Symbol;Acc:HGNC:14928]	0.33	0.00
612	ENSG00000087586	AURKA	aurora kinase A [Source:HGNC Symbol;Acc:HGNC:11393]	0.33	0.00
613	ENSG00000196646	ZNF136	zinc finger protein 136 [Source:HGNC Symbol;Acc:HGNC:12920]	0.33	0.00
614	ENSG00000145247	OCIAD2	OCIA domain containing 2 [Source:HGNC Symbol;Acc:HGNC:28685]	0.33	0.00
615	ENSG00000114450	GNB4	guanine nucleotide binding protein (G protein), beta polypeptide 4 [Source:HGNC Symbol;Acc:HGNC:20731]	0.33	0.00
616	ENSG00000183077	AFMID	arylformamidase [Source:HGNC Symbol;Acc:HGNC:20910]	0.33	0.00
617	ENSG00000166130	IKBIP	IKBKB interacting protein [Source:HGNC Symbol;Acc:HGNC:26430]	0.34	0.00
618	ENSG00000130340	SNX9	sorting nexin 9 [Source:HGNC Symbol;Acc:HGNC:14973]	0.34	0.00
619	ENSG00000175581	MRPL48	mitochondrial ribosomal protein L48 [Source:HGNC Symbol;Acc:HGNC:16653]	0.34	0.00
620	ENSG00000110811	P3H3	prolyl 3-hydroxylase 3 [Source:HGNC Symbol;Acc:HGNC:19318]	0.34	0.00
621	ENSG00000213626	LBH	limb bud and heart development [Source:HGNC Symbol;Acc:HGNC:29532]	0.34	0.00
622	ENSG00000145390	USP53	ubiquitin specific peptidase 53 [Source:HGNC Symbol;Acc:HGNC:29255]	0.34	0.00
623	ENSG00000171604	CXXC5	CXXC finger protein 5 [Source:HGNC Symbol;Acc:HGNC:26943]	0.35	0.00
624	ENSG00000122644	ARL4A	ADP ribosylation factor like GTPase 4A [Source:HGNC Symbol;Acc:HGNC:695]	0.35	0.00
625	ENSG00000163166	IWS1	IWS1 homolog (S. cerevisiae) [Source:HGNC Symbol;Acc:HGNC:25467]	0.35	0.00
626	ENSG00000122861	PLAU	plasminogen activator, urokinase [Source:HGNC Symbol;Acc:HGNC:9052]	0.35	0.00
627	ENSG00000152818	UTRN	utrophin [Source:HGNC Symbol;Acc:HGNC:12635]	0.35	0.00
628	ENSG00000167565	SERTAD3	SERTA domain containing 3 [Source:HGNC Symbol;Acc:HGNC:17931]	0.35	0.00
629	ENSG00000102316	MAGED2	MAGE family member D2 [Source:HGNC Symbol;Acc:HGNC:16353]	0.35	0.00
630	ENSG00000100154	TTC28	tetratricopeptide repeat domain 28 [Source:HGNC Symbol;Acc:HGNC:29179]	0.35	0.00
631	ENSG00000134247	PTGFRN	prostaglandin F2 receptor inhibitor [Source:HGNC Symbol;Acc:HGNC:9601]	0.35	0.00
632	ENSG00000197905	TEAD4	TEA domain family member 4 [Source:HGNC Symbol;Acc:HGNC:11717]	0.35	0.00
633	ENSG00000132434	LANCL2	LanC lantibiotic synthetase component C-like 2 (bacterial) [Source:HGNC Symbol;Acc:HGNC:6509]	0.36	0.00
634	ENSG00000124588	NQO2	NAD(P)H dehydrogenase, quinone 2 [Source:HGNC Symbol;Acc:HGNC:7856]	0.36	0.00
635	ENSG00000102349	KLF8	Kruppel-like factor 8 [Source:HGNC Symbol;Acc:HGNC:6351]	0.36	0.00
636	ENSG00000198915	RASGEF1A	RasGEF domain family member 1A [Source:HGNC Symbol;Acc:HGNC:24246]	0.36	0.00
637	ENSG00000182621	PLCB1	phospholipase C beta 1 [Source:HGNC Symbol;Acc:HGNC:15917]	0.36	0.00
638	ENSG00000196683	TOMM7	translocase of outer mitochondrial membrane 7 homolog (yeast) [Source:HGNC Symbol;Acc:HGNC:21648]	0.36	0.00
639	ENSG00000197608	ZNF841	zinc finger protein 841 [Source:HGNC Symbol;Acc:HGNC:27611]	0.36	0.00

640	ENSG00000154122	ANKH	ANKH inorganic pyrophosphate transport regulator [Source:HGNC Symbol;Acc:HGNC:15492]	0.36	0.00
641	ENSG00000144677	CTDSPL	CTD small phosphatase like [Source:HGNC Symbol;Acc:HGNC:16890]	0.36	0.00
642	ENSG00000120686	UFM1	ubiquitin-fold modifier 1 [Source:HGNC Symbol;Acc:HGNC:20597]	0.37	0.00
643	ENSG00000253276	CCDC71L	coiled-coil domain containing 71-like [Source:HGNC Symbol;Acc:HGNC:26685]	0.37	0.00
644	ENSG00000151136	BTBD11	BTB (POZ) domain containing 11 [Source:HGNC Symbol;Acc:HGNC:23844]	0.37	0.00
645	ENSG00000100353	EIF3D	eukaryotic translation initiation factor 3 subunit D [Source:HGNC Symbol;Acc:HGNC:3278]	0.37	0.00
646	ENSG00000121005	CRISPLD1	cysteine-rich secretory protein LCCL domain containing 1 [Source:HGNC Symbol;Acc:HGNC:18206]	0.37	0.00
647	ENSG00000101004	NINL	ninein like [Source:HGNC Symbol;Acc:HGNC:29163]	0.37	0.00
648	ENSG00000198408	MGEA5	meningioma expressed antigen 5 (hyaluronidase) [Source:HGNC Symbol;Acc:HGNC:7056]	0.38	0.00
649	ENSG00000143333	RGS16	regulator of G-protein signaling 16 [Source:HGNC Symbol;Acc:HGNC:9997]	0.38	0.00
650	ENSG00000154040	CABYR	calcium binding tyrosine-(Y)-phosphorylation regulated [Source:HGNC Symbol;Acc:HGNC:15569]	0.38	0.00
651	ENSG00000182704	TSKU	tsukushi, small leucine rich proteoglycan [Source:HGNC Symbol;Acc:HGNC:28850]	0.38	0.00
652	ENSG00000071205	ARHGAP10	Rho GTPase activating protein 10 [Source:HGNC Symbol;Acc:HGNC:26099]	0.38	0.00
653	ENSG00000172456	FGGY	FGGY carbohydrate kinase domain containing [Source:HGNC Symbol;Acc:HGNC:25610]	0.38	0.00
654	ENSG00000151388	ADAMTS12	ADAM metallopeptidase with thrombospondin type 1 motif 12 [Source:HGNC Symbol;Acc:HGNC:14605]	0.39	0.00
655	ENSG00000007384	RHBDF1	rhomboid 5 homolog 1 (Drosophila) [Source:HGNC Symbol;Acc:HGNC:20561]	0.39	0.00
656	ENSG00000116678	LEPR	leptin receptor [Source:HGNC Symbol;Acc:HGNC:6554]	0.39	0.00
657	ENSG00000198938	MT-CO3	mitochondrially encoded cytochrome c oxidase III [Source:HGNC Symbol;Acc:HGNC:7422]	0.39	0.00
658	ENSG00000163520	FBLN2	fibulin 2 [Source:HGNC Symbol;Acc:HGNC:3601]	0.39	0.00
659	ENSG00000185340	GAS2L1	growth arrest specific 2 like 1 [Source:HGNC Symbol;Acc:HGNC:16955]	0.39	0.00
660	ENSG00000169093	ASMTL	acetylserotonin O-methyltransferase-like [Source:HGNC Symbol;Acc:HGNC:751]	0.39	0.00
661	ENSG00000136826	KLF4	Kruppel-like factor 4 (gut) [Source:HGNC Symbol;Acc:HGNC:6348]	0.39	0.00
662	ENSG00000186468	RPS23	ribosomal protein S23 [Source:HGNC Symbol;Acc:HGNC:10410]	0.39	0.00
663	ENSG00000135333	EPHA7	EPH receptor A7 [Source:HGNC Symbol;Acc:HGNC:3390]	0.40	0.00
664	ENSG00000105193	RPS16	ribosomal protein S16 [Source:HGNC Symbol;Acc:HGNC:10396]	0.40	0.00
665	ENSG00000137193	PIM1	Pim-1 proto-oncogene, serine/threonine kinase [Source:HGNC Symbol;Acc:HGNC:8986]	0.40	0.00
666	ENSG00000137955	RABGGTB	Rab geranylgeranyltransferase, beta subunit [Source:HGNC Symbol;Acc:HGNC:9796]	0.40	0.00
667	ENSG00000080546	SESN1	sestrin 1 [Source:HGNC Symbol;Acc:HGNC:21595]	0.40	0.00
668	ENSG00000136732	GYPC	glycophorin C (Gerbich blood group) [Source:HGNC Symbol;Acc:HGNC:4704]	0.40	0.00
669	ENSG00000204387	C6orf48	chromosome 6 open reading frame 48 [Source:HGNC Symbol;Acc:HGNC:19078]	0.40	0.00
670	ENSG00000150593	PDCD4	programmed cell death 4 (neoplastic transformation inhibitor) [Source:HGNC Symbol;Acc:HGNC:8763]	0.40	0.00
671	ENSG00000114948	ADAM23	ADAM metallopeptidase domain 23 [Source:HGNC Symbol;Acc:HGNC:202]	0.41	0.00
672	ENSG00000076685	NT5C2	5'-nucleotidase, cytosolic II [Source:HGNC Symbol;Acc:HGNC:8022]	0.41	0.00
673	ENSG00000107518	ATRNL1	attractin like 1 [Source:HGNC Symbol;Acc:HGNC:29063]	0.41	0.00

674	ENSG00000146242	TPBG	trophoblast glycoprotein [Source:HGNC Symbol;Acc:HGNC:12004]	0.41	0.00
675	ENSG00000196352	CD55	CD55 molecule, decay accelerating factor for complement (Cromer blood group) [Source:HGNC Symbol;Acc:HGNC:2665]	0.41	0.00
676	ENSG00000175215	CTDSP2	CTD small phosphatase 2 [Source:HGNC Symbol;Acc:HGNC:17077]	0.41	0.00
677	ENSG00000006194	ZNF263	zinc finger protein 263 [Source:HGNC Symbol;Acc:HGNC:13056]	0.41	0.00
678	ENSG00000100916	BRMS1L	breast cancer metastasis-suppressor 1-like [Source:HGNC Symbol;Acc:HGNC:20512]	0.41	0.00
679	ENSG00000134324	LPIN1	lipin 1 [Source:HGNC Symbol;Acc:HGNC:13345]	0.41	0.00
680	ENSG00000166025	AMOTL1	angiomotin like 1 [Source:HGNC Symbol;Acc:HGNC:17811]	0.41	0.00
681	ENSG00000077044	DGKD	diacylglycerol kinase delta [Source:HGNC Symbol;Acc:HGNC:2851]	0.42	0.00
682	ENSG00000158710	TAGLN2	transgelin 2 [Source:HGNC Symbol;Acc:HGNC:11554]	0.42	0.00
683	ENSG00000172197	MBOAT1	membrane bound O-acyltransferase domain containing 1 [Source:HGNC Symbol;Acc:HGNC:21579]	0.42	0.00
684	ENSG00000112559	MDFI	MyoD family inhibitor [Source:HGNC Symbol;Acc:HGNC:6967]	0.42	0.00
685	ENSG00000136010	ALDH1L2	aldehyde dehydrogenase 1 family member L2 [Source:HGNC Symbol;Acc:HGNC:26777]	0.42	0.00
686	ENSG00000128335	APOL2	apolipoprotein L2 [Source:HGNC Symbol;Acc:HGNC:619]	0.42	0.00
687	ENSG00000176204	LRRTM4	leucine rich repeat transmembrane neuronal 4 [Source:HGNC Symbol;Acc:HGNC:19411]	0.43	0.00
688	ENSG00000122863	CHST3	carbohydrate (chondroitin 6) sulfotransferase 3 [Source:HGNC Symbol;Acc:HGNC:1971]	0.43	0.00
689	ENSG00000134574	DDB2	damage-specific DNA binding protein 2 [Source:HGNC Symbol;Acc:HGNC:2718]	0.43	0.00
690	ENSG00000146409	SLC18B1	solute carrier family 18, subfamily B, member 1 [Source:HGNC Symbol;Acc:HGNC:21573]	0.43	0.00
691	ENSG00000148426	PROSER2	proline and serine rich 2 [Source:HGNC Symbol;Acc:HGNC:23728]	0.43	0.00
692	ENSG00000167981	ZNF597	zinc finger protein 597 [Source:HGNC Symbol;Acc:HGNC:26573]	0.43	0.00
693	ENSG00000170899	GSTA4	glutathione S-transferase alpha 4 [Source:HGNC Symbol;Acc:HGNC:4629]	0.43	0.00
694	ENSG00000164932	CTHRC1	collagen triple helix repeat containing 1 [Source:HGNC Symbol;Acc:HGNC:18831]	0.44	0.00
695	ENSG00000156140	ADAMTS3	ADAM metallopeptidase with thrombospondin type 1 motif 3 [Source:HGNC Symbol;Acc:HGNC:219]	0.44	0.00
696	ENSG00000048052	HDAC9	histone deacetylase 9 [Source:HGNC Symbol;Acc:HGNC:14065]	0.44	0.00
697	ENSG00000130066	SAT1	spermidine/spermine N1-acetyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:10540]	0.44	0.00
698	ENSG00000120833	SOCS2	suppressor of cytokine signaling 2 [Source:HGNC Symbol;Acc:HGNC:19382]	0.44	0.00
699	ENSG00000197324	LRP10	LDL receptor related protein 10 [Source:HGNC Symbol;Acc:HGNC:14553]	0.44	0.00
700	ENSG00000111057	KRT18	keratin 18, type I [Source:HGNC Symbol;Acc:HGNC:6430]	0.44	0.00
701	ENSG00000185088	RPS27L	ribosomal protein S27 like [Source:HGNC Symbol;Acc:HGNC:18476]	0.44	0.00
702	ENSG00000156869	FRRS1	ferric-chelate reductase 1 [Source:HGNC Symbol;Acc:HGNC:27622]	0.45	0.00
703	ENSG00000067177	PHKA1	phosphorylase kinase, alpha 1 (muscle) [Source:HGNC Symbol;Acc:HGNC:8925]	0.45	0.00
704	ENSG00000105327	BBC3	BCL2 binding component 3 [Source:HGNC Symbol;Acc:HGNC:17868]	0.45	0.00
705	ENSG00000183696	UPP1	uridine phosphorylase 1 [Source:HGNC Symbol;Acc:HGNC:12576]	0.45	0.00
706	ENSG00000127955	GNAI1	guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 1 [Source:HGNC Symbol;Acc:HGNC:4384]	0.45	0.00
707	ENSG00000080573	COL5A3	collagen, type V, alpha 3 [Source:HGNC Symbol;Acc:HGNC:14864]	0.45	0.00
708	ENSG00000089472	HEPH	hephaestin [Source:HGNC Symbol;Acc:HGNC:4866]	0.45	0.00
709	ENSG00000257315	ZBED6	zinc finger, BED-type containing 6 [Source:HGNC Symbol;Acc:HGNC:33273]	0.45	0.00

710	ENSG00000136098	NEK3	NIMA-related kinase 3 [Source:HGNC Symbol;Acc:HGNC:7746]	0.46	0.00
711	ENSG00000072422	RHOBTB1	Rho-related BTB domain containing 1 [Source:HGNC Symbol;Acc:HGNC:18738]	0.46	0.00
712	ENSG00000136156	ITM2B	integral membrane protein 2B [Source:HGNC Symbol;Acc:HGNC:6174]	0.46	0.00
713	ENSG00000170421	KRT8	keratin 8, type II [Source:HGNC Symbol;Acc:HGNC:6446]	0.47	0.00
714	ENSG00000115414	FN1	fibronectin 1 [Source:HGNC Symbol;Acc:HGNC:3778]	0.47	0.00
715	ENSG00000177311	ZBTB38	zinc finger and BTB domain containing 38 [Source:HGNC Symbol;Acc:HGNC:26636]	0.47	0.00
716	ENSG00000164687	FABP5	fatty acid binding protein 5 (psoriasis-associated) [Source:HGNC Symbol;Acc:HGNC:3560]	0.47	0.00
717	ENSG00000197702	PARVA	parvin alpha [Source:HGNC Symbol;Acc:HGNC:14652]	0.47	0.00
718	ENSG00000145536	ADAMTS16	ADAM metallopeptidase with thrombospondin type 1 motif 16 [Source:HGNC Symbol;Acc:HGNC:17108]	0.47	0.00
719	ENSG00000179046	TRIML2	tripartite motif family like 2 [Source:HGNC Symbol;Acc:HGNC:26378]	0.47	0.00
720	ENSG00000115112	TFCP2L1	transcription factor CP2-like 1 [Source:HGNC Symbol;Acc:HGNC:17925]	0.48	0.00
721	ENSG00000101850	GPR143	G protein-coupled receptor 143 [Source:HGNC Symbol;Acc:HGNC:20145]	0.48	0.00
722	ENSG00000160145	KALRN	kalirin, RhoGEF kinase [Source:HGNC Symbol;Acc:HGNC:4814]	0.48	0.00
723	ENSG00000082397	EPB41L3	erythrocyte membrane protein band 4.1-like 3 [Source:HGNC Symbol;Acc:HGNC:3380]	0.48	0.00
724	ENSG00000166250	CLMP	CXADR-like membrane protein [Source:HGNC Symbol;Acc:HGNC:24039]	0.48	0.00
725	ENSG00000069869	NEDD4	neural precursor cell expressed, developmentally down-regulated 4, E3 ubiquitin protein ligase [Source:HGNC Symbol;Acc:HGNC:7727]	0.48	0.00
726	ENSG00000242125	SNHG3	small nucleolar RNA host gene 3 [Source:HGNC Symbol;Acc:HGNC:10118]	0.48	0.00
727	ENSG00000154783	FGD5	FYVE, RhoGEF and PH domain containing 5 [Source:HGNC Symbol;Acc:HGNC:19117]	0.49	0.00
728	ENSG00000133121	STARD13	StAR related lipid transfer domain containing 13 [Source:HGNC Symbol;Acc:HGNC:19164]	0.49	0.00
729	ENSG00000112984	KIF20A	kinesin family member 20A [Source:HGNC Symbol;Acc:HGNC:9787]	0.49	0.00
730	ENSG00000116741	RGS2	regulator of G-protein signaling 2 [Source:HGNC Symbol;Acc:HGNC:9998]	0.49	0.00
731	ENSG00000027075	PRKCH	protein kinase C, eta [Source:HGNC Symbol;Acc:HGNC:9403]	0.49	0.00
732	ENSG00000137672	TRPC6	transient receptor potential cation channel, subfamily C, member 6 [Source:HGNC Symbol;Acc:HGNC:12338]	0.49	0.00
733	ENSG00000163359	COL6A3	collagen, type VI, alpha 3 [Source:HGNC Symbol;Acc:HGNC:2213]	0.49	0.00
734	ENSG00000173559	NABP1	nucleic acid binding protein 1 [Source:HGNC Symbol;Acc:HGNC:26232]	0.49	0.00
735	ENSG00000169570	DTWD2	DTW domain containing 2 [Source:HGNC Symbol;Acc:HGNC:19334]	0.49	0.00
736	ENSG00000123131	PRDX4	peroxiredoxin 4 [Source:HGNC Symbol;Acc:HGNC:17169]	0.49	0.00
737	ENSG00000228526	MIR34AHG	MIR34A host gene [Source:HGNC Symbol;Acc:HGNC:51913]	0.49	0.00
738	ENSG00000026508	CD44	CD44 molecule (Indian blood group) [Source:HGNC Symbol;Acc:HGNC:1681]	0.50	0.00
739	ENSG00000139117	CPNE8	copine VIII [Source:HGNC Symbol;Acc:HGNC:23498]	0.50	0.00
740	ENSG00000184384	MAML2	mastermind like transcriptional coactivator 2 [Source:HGNC Symbol;Acc:HGNC:16259]	0.50	0.00
741	ENSG00000101665	SMAD7	SMAD family member 7 [Source:HGNC Symbol;Acc:HGNC:6773]	0.51	0.00
742	ENSG00000173482	PTPRM	protein tyrosine phosphatase, receptor type, M [Source:HGNC Symbol;Acc:HGNC:9675]	0.51	0.00
743	ENSG00000128283	CDC42EP1	CDC42 effector protein (Rho GTPase binding) 1 [Source:HGNC Symbol;Acc:HGNC:17014]	0.51	0.00
744	ENSG00000115350	POLE4	polymerase (DNA-directed), epsilon 4, accessory subunit [Source:HGNC Symbol;Acc:HGNC:18755]	0.51	0.00

745	ENSG00000145681	HAPLN1	hyaluronan and proteoglycan link protein 1 [Source:HGNC Symbol;Acc:HGNC:2380]	0.51	0.00
746	ENSG00000107104	KANK1	KN motif and ankyrin repeat domains 1 [Source:HGNC Symbol;Acc:HGNC:19309]	0.51	0.00
747	ENSG00000161513	FDXR	ferredoxin reductase [Source:HGNC Symbol;Acc:HGNC:3642]	0.52	0.00
748	ENSG00000127241	MASP1	mannan-binding lectin serine peptidase 1 (C4/C2 activating component of Ra-reactive factor) [Source:HGNC Symbol;Acc:HGNC:6901]	0.52	0.00
749	ENSG00000151892	GFRA1	GDNF family receptor alpha 1 [Source:HGNC Symbol;Acc:HGNC:4243]	0.52	0.00
750	ENSG00000112343	TRIM38	tripartite motif containing 38 [Source:HGNC Symbol;Acc:HGNC:10059]	0.52	0.00
751	ENSG00000113448	PDE4D	phosphodiesterase 4D [Source:HGNC Symbol;Acc:HGNC:8783]	0.52	0.00
752	ENSG00000134802	SLC43A3	solute carrier family 43 member 3 [Source:HGNC Symbol;Acc:HGNC:17466]	0.52	0.00
753	ENSG00000120549	KIAA1217	KIAA1217 [Source:HGNC Symbol;Acc:HGNC:25428]	0.52	0.00
754	ENSG00000112186	CAP2	CAP, adenylate cyclase-associated protein, 2 (yeast) [Source:HGNC Symbol;Acc:HGNC:20039]	0.52	0.00
755	ENSG00000115380	EFEMP1	EGF containing fibulin-like extracellular matrix protein 1 [Source:HGNC Symbol;Acc:HGNC:3218]	0.52	0.00
756	ENSG00000171055	FEZ2	fasciculation and elongation protein zeta 2 (zygin II) [Source:HGNC Symbol;Acc:HGNC:3660]	0.52	0.00
757	ENSG00000139926	FRMD6	FERM domain containing 6 [Source:HGNC Symbol;Acc:HGNC:19839]	0.53	0.00
758	ENSG00000010539	ZNF200	zinc finger protein 200 [Source:HGNC Symbol;Acc:HGNC:12993]	0.53	0.00
759	ENSG00000050767	COL23A1	collagen, type XXIII, alpha 1 [Source:HGNC Symbol;Acc:HGNC:22990]	0.53	0.00
760	ENSG00000136960	ENPP2	ectonucleotide pyrophosphatase/phosphodiesterase 2 [Source:HGNC Symbol;Acc:HGNC:3357]	0.53	0.00
761	ENSG00000122507	BBS9	Bardet-Biedl syndrome 9 [Source:HGNC Symbol;Acc:HGNC:30000]	0.53	0.00
762	ENSG00000132274	TRIM22	tripartite motif containing 22 [Source:HGNC Symbol;Acc:HGNC:16379]	0.53	0.00
763	ENSG00000115828	QPCT	glutaminyl-peptide cyclotransferase [Source:HGNC Symbol;Acc:HGNC:9753]	0.53	0.00
764	ENSG00000103740	ACSBG1	acyl-CoA synthetase bubblegum family member 1 [Source:HGNC Symbol;Acc:HGNC:29567]	0.54	0.00
765	ENSG00000175832	ETV4	ets variant 4 [Source:HGNC Symbol;Acc:HGNC:3493]	0.54	0.00
766	ENSG00000060140	STYK1	serine/threonine/tyrosine kinase 1 [Source:HGNC Symbol;Acc:HGNC:18889]	0.54	0.00
767	ENSG00000150394	CDH8	cadherin 8, type 2 [Source:HGNC Symbol;Acc:HGNC:1767]	0.54	0.00
768	ENSG00000169116	PARM1	prostate androgen-regulated mucin-like protein 1 [Source:HGNC Symbol;Acc:HGNC:24536]	0.54	0.00
769	ENSG00000189184	PCDH18	protocadherin 18 [Source:HGNC Symbol;Acc:HGNC:14268]	0.54	0.00
770	ENSG00000152377	SPOCK1	sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 1 [Source:HGNC Symbol;Acc:HGNC:11251]	0.54	0.00
771	ENSG00000135917	SLC19A3	solute carrier family 19 (thiamine transporter), member 3 [Source:HGNC Symbol;Acc:HGNC:16266]	0.55	0.00
772	ENSG00000179542	SLITRK4	SLIT and NTRK like family member 4 [Source:HGNC Symbol;Acc:HGNC:23502]	0.55	0.00
773	ENSG00000075651	PLD1	phospholipase D1 [Source:HGNC Symbol;Acc:HGNC:9067]	0.55	0.00
774	ENSG00000154734	ADAMTS1	ADAM metallopeptidase with thrombospondin type 1 motif 1 [Source:HGNC Symbol;Acc:HGNC:217]	0.55	0.00
775	ENSG00000082781	ITGB5	integrin subunit beta 5 [Source:HGNC Symbol;Acc:HGNC:6160]	0.55	0.00
776	ENSG00000173267	SNCG	synuclein gamma [Source:HGNC Symbol;Acc:HGNC:11141]	0.55	0.00
777	ENSG00000053747	LAMA3	laminin subunit alpha 3 [Source:HGNC Symbol;Acc:HGNC:6483]	0.56	0.00
778	ENSG00000104611	SH2D4A	SH2 domain containing 4A [Source:HGNC Symbol;Acc:HGNC:26102]	0.56	0.00
779	ENSG00000104419	NDRG1	N-myc downstream regulated 1 [Source:HGNC Symbol;Acc:HGNC:7679]	0.56	0.00

780	ENSG00000007237	GAS7	growth arrest specific 7 [Source:HGNC Symbol;Acc:HGNC:4169]	0.56	0.00
781	ENSG00000115290	GRB14	growth factor receptor bound protein 14 [Source:HGNC Symbol;Acc:HGNC:4565]	0.56	0.00
782	ENSG00000101938	CHRDL1	chordin-like 1 [Source:HGNC Symbol;Acc:HGNC:29861]	0.56	0.00
783	ENSG00000100031	GGT1	gamma-glutamyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:4250]	0.56	0.00
784	ENSG00000100031	GGT1	gamma-glutamyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:4250]	0.56	0.00
785	ENSG00000100031	GGT1	gamma-glutamyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:4250]	0.56	0.00
786	ENSG00000100031	GGT1	gamma-glutamyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:4250]	0.56	0.00
787	ENSG00000100031	GGT1	gamma-glutamyltransferase 1 [Source:HGNC Symbol;Acc:HGNC:4250]	0.56	0.00
788	ENSG00000164292	RHOBTB3	Rho-related BTB domain containing 3 [Source:HGNC Symbol;Acc:HGNC:18757]	0.57	0.00
789	ENSG00000112414	ADGRG6	adhesion G protein-coupled receptor G6 [Source:HGNC Symbol;Acc:HGNC:13841]	0.57	0.00
790	ENSG00000145431	PDGFC	platelet derived growth factor C [Source:HGNC Symbol;Acc:HGNC:8801]	0.58	0.00
791	ENSG00000158125	XDH	xanthine dehydrogenase [Source:HGNC Symbol;Acc:HGNC:12805]	0.58	0.00
792	ENSG00000164070	HSPA4L	heat shock protein family A (Hsp70) member 4 like [Source:HGNC Symbol;Acc:HGNC:17041]	0.58	0.00
793	ENSG00000158828	PINK1	PTEN induced putative kinase 1 [Source:HGNC Symbol;Acc:HGNC:14581]	0.59	0.00
794	ENSG00000172380	GNG12	guanine nucleotide binding protein (G protein), gamma 12 [Source:HGNC Symbol;Acc:HGNC:19663]	0.59	0.00
795	ENSG00000187676	B3GLCT	beta 3-glucosyltransferase [Source:HGNC Symbol;Acc:HGNC:20207]	0.59	0.00
796	ENSG00000158089	GALNT14	polypeptide N-acetylgalactosaminyltransferase 14 [Source:HGNC Symbol;Acc:HGNC:22946]	0.59	0.00
797	ENSG00000047617	ANO2	anoctamin 2, calcium activated chloride channel [Source:HGNC Symbol;Acc:HGNC:1183]	0.59	0.00
798	ENSG00000117069	ST6GALNAC5	ST6 (alpha-N-acetyl-neuraminyl-2,3-beta-galactosyl-1,3)-N-acetylgalactosaminide alpha-2,6-sialyltransferase 5 [Source:HGNC Symbol;Acc:HGNC:19342]	0.59	0.00
799	ENSG00000103196	CRISPLD2	cysteine-rich secretory protein LCCL domain containing 2 [Source:HGNC Symbol;Acc:HGNC:25248]	0.59	0.00
800	ENSG00000120708	TGFBI	transforming growth factor beta induced [Source:HGNC Symbol;Acc:HGNC:11771]	0.59	0.00
801	ENSG00000140092	FBLN5	fibulin 5 [Source:HGNC Symbol;Acc:HGNC:3602]	0.60	0.00
802	ENSG00000130830	MPP1	membrane protein, palmitoylated 1 [Source:HGNC Symbol;Acc:HGNC:7219]	0.60	0.00
803	ENSG00000104324	CPQ	carboxypeptidase Q [Source:HGNC Symbol;Acc:HGNC:16910]	0.60	0.00
804	ENSG00000178538	CA8	carbonic anhydrase VIII [Source:HGNC Symbol;Acc:HGNC:1382]	0.60	0.00
805	ENSG00000172986	GXYLT2	glucoside xylosyltransferase 2 [Source:HGNC Symbol;Acc:HGNC:33383]	0.60	0.00
806	ENSG00000147465	STAR	steroidogenic acute regulatory protein [Source:HGNC Symbol;Acc:HGNC:11359]	0.61	0.00
807	ENSG00000131080	EDA2R	ectodysplasin A2 receptor [Source:HGNC Symbol;Acc:HGNC:17756]	0.61	0.00
808	ENSG00000150722	PPP1R1C	protein phosphatase 1 regulatory inhibitor subunit 1C [Source:HGNC Symbol;Acc:HGNC:14940]	0.61	0.00
809	ENSG00000074527	NTN4	netrin 4 [Source:HGNC Symbol;Acc:HGNC:13658]	0.61	0.00
810	ENSG00000163219	ARHGAP25	Rho GTPase activating protein 25 [Source:HGNC Symbol;Acc:HGNC:28951]	0.61	0.00
811	ENSG00000187475	HIST1H1T	histone cluster 1, H1t [Source:HGNC Symbol;Acc:HGNC:4720]	0.62	0.00
812	ENSG00000134762	DSC3	desmocollin 3 [Source:HGNC Symbol;Acc:HGNC:3037]	0.62	0.00
813	ENSG00000118523	CTGF	connective tissue growth factor [Source:HGNC Symbol;Acc:HGNC:2500]	0.62	0.00
814	ENSG00000188984	AADACL3	arylacetamide deacetylase-like 3 [Source:HGNC Symbol;Acc:HGNC:32037]	0.62	0.00
815	ENSG00000109452	INPP4B	inositol polyphosphate-4-phosphatase type II B [Source:HGNC Symbol;Acc:HGNC:6075]	0.62	0.00
816	ENSG00000124762	CDKN1A	cyclin-dependent kinase inhibitor 1A (p21, Cip1) [Source:HGNC Symbol;Acc:HGNC:1784]	0.63	0.00

817	ENSG00000138821	SLC39A8	solute carrier family 39 (zinc transporter), member 8 [Source:HGNC Symbol;Acc:HGNC:20862]	0.63	0.00
818	ENSG00000134954	ETS1	v-ets avian erythroblastosis virus E26 oncogene homolog 1 [Source:HGNC Symbol;Acc:HGNC:3488]	0.63	0.00
819	ENSG00000188517	COL25A1	collagen, type XXV, alpha 1 [Source:HGNC Symbol;Acc:HGNC:18603]	0.63	0.00
820	ENSG00000165591	FAAH2	fatty acid amide hydrolase 2 [Source:HGNC Symbol;Acc:HGNC:26440]	0.64	0.00
821	ENSG00000102547	CAB39L	calcium binding protein 39 like [Source:HGNC Symbol;Acc:HGNC:20290]	0.64	0.00
822	ENSG00000166736	HTR3A	5-hydroxytryptamine (serotonin) receptor 3A, ionotropic [Source:HGNC Symbol;Acc:HGNC:5297]	0.65	0.00
823	ENSG00000091428	RAPGEF4	Rap guanine nucleotide exchange factor 4 [Source:HGNC Symbol;Acc:HGNC:16626]	0.65	0.00
824	ENSG00000164237	CMBL	carboxymethylenebutenolidase homolog (Pseudomonas) [Source:HGNC Symbol;Acc:HGNC:25090]	0.66	0.00
825	ENSG00000151617	EDNRA	endothelin receptor type A [Source:HGNC Symbol;Acc:HGNC:3179]	0.66	0.00
826	ENSG00000177398	UMODL1	uromodulin like 1 [Source:HGNC Symbol;Acc:HGNC:12560]	0.67	0.00
827	ENSG00000258545	RHOXF1-AS1	RHOXF1 antisense RNA 1 [Source:HGNC Symbol;Acc:HGNC:51582]	0.67	0.00
828	ENSG00000163071	SPATA18	spermatogenesis associated 18 [Source:HGNC Symbol;Acc:HGNC:29579]	0.67	0.00
829	ENSG00000183508	FAM46C	family with sequence similarity 46 member C [Source:HGNC Symbol;Acc:HGNC:24712]	0.68	0.00
830	ENSG00000134042	MRO	maestro [Source:HGNC Symbol;Acc:HGNC:24121]	0.68	0.00
831	ENSG00000164136	IL15	interleukin 15 [Source:HGNC Symbol;Acc:HGNC:5977]	0.69	0.00
832	ENSG00000127084	FGD3	FYVE, RhoGEF and PH domain containing 3 [Source:HGNC Symbol;Acc:HGNC:16027]	0.69	0.00
833	ENSG00000150540	HNMT	histamine N-methyltransferase [Source:HGNC Symbol;Acc:HGNC:5028]	0.69	0.00
834	ENSG00000175445	LPL	lipoprotein lipase [Source:HGNC Symbol;Acc:HGNC:6677]	0.69	0.00
835	ENSG00000135298	ADGRB3	adhesion G protein-coupled receptor B3 [Source:HGNC Symbol;Acc:HGNC:945]	0.69	0.00
836	ENSG00000153071	DAB2	Dab, mitogen-responsive phosphoprotein, homolog 2 (Drosophila) [Source:HGNC Symbol;Acc:HGNC:2662]	0.70	0.00
837	ENSG00000183230	CTNNA3	catenin alpha 3 [Source:HGNC Symbol;Acc:HGNC:2511]	0.70	0.00
838	ENSG00000049130	KITLG	KIT ligand [Source:HGNC Symbol;Acc:HGNC:6343]	0.71	0.00
839	ENSG00000164946	FREM1	FRAS1 related extracellular matrix 1 [Source:HGNC Symbol;Acc:HGNC:23399]	0.71	0.00
840	ENSG00000205730	ITPRIPL2	inositol 1,4,5-trisphosphate receptor interacting protein-like 2 [Source:HGNC Symbol;Acc:HGNC:27257]	0.71	0.00
841	ENSG00000146592	CREB5	cAMP responsive element binding protein 5 [Source:HGNC Symbol;Acc:HGNC:16844]	0.71	0.00
842	ENSG00000049192	ADAMTS6	ADAM metallopeptidase with thrombospondin type 1 motif 6 [Source:HGNC Symbol;Acc:HGNC:222]	0.71	0.00
843	ENSG00000110799	VWF	von Willebrand factor [Source:HGNC Symbol;Acc:HGNC:12726]	0.72	0.00
844	ENSG00000168772	CXXC4	CXXC finger protein 4 [Source:HGNC Symbol;Acc:HGNC:24593]	0.72	0.00
845	ENSG00000118785	SPP1	secreted phosphoprotein 1 [Source:HGNC Symbol;Acc:HGNC:11255]	0.72	0.00
846	ENSG00000198756	COLGALT2	collagen beta(1-O)galactosyltransferase 2 [Source:HGNC Symbol;Acc:HGNC:16790]	0.72	0.00
847	ENSG00000174564	IL20RB	interleukin 20 receptor subunit beta [Source:HGNC Symbol;Acc:HGNC:6004]	0.72	0.00
848	ENSG00000182645	CCDC172	coiled-coil domain containing 172 [Source:HGNC Symbol;Acc:HGNC:30524]	0.73	0.00
849	ENSG00000143318	CASQ1	calsequestrin 1 [Source:HGNC Symbol;Acc:HGNC:1512]	0.74	0.00
850	ENSG00000138772	ANXA3	annexin A3 [Source:HGNC Symbol;Acc:HGNC:541]	0.74	0.00
851	ENSG00000069535	MAOB	monoamine oxidase B [Source:HGNC Symbol;Acc:HGNC:6834]	0.75	0.00

852	ENSG00000133135	RNF128	ring finger protein 128, E3 ubiquitin protein ligase [Source:HGNC Symbol;Acc:HGNC:21153]	0.75	0.00
853	ENSG00000180801	ARSJ	arylsulfatase family member J [Source:HGNC Symbol;Acc:HGNC:26286]	0.76	0.00
854	ENSG00000075213	SEMA3A	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3A [Source:HGNC Symbol;Acc:HGNC:10723]	0.76	0.00
855	ENSG00000182326	C1S	complement component 1, s subcomponent [Source:HGNC Symbol;Acc:HGNC:1247]	0.77	0.00
856	ENSG00000155511	GRIA1	glutamate receptor, ionotropic, AMPA 1 [Source:HGNC Symbol;Acc:HGNC:4571]	0.78	0.00
857	ENSG00000115468	EFHD1	EF-hand domain family member D1 [Source:HGNC Symbol;Acc:HGNC:29556]	0.79	0.00
858	ENSG00000107954	NEURL1	neuralized E3 ubiquitin protein ligase 1 [Source:HGNC Symbol;Acc:HGNC:7761]	0.79	0.00
859	ENSG00000204262	COL5A2	collagen, type V, alpha 2 [Source:HGNC Symbol;Acc:HGNC:2210]	0.79	0.00
860	ENSG00000091986	CCDC80	coiled-coil domain containing 80 [Source:HGNC Symbol;Acc:HGNC:30649]	0.80	0.00
861	ENSG00000108375	RNF43	ring finger protein 43 [Source:HGNC Symbol;Acc:HGNC:18505]	0.80	0.00
862	ENSG00000123700	KCNJ2	potassium channel, inwardly rectifying subfamily J, member 2 [Source:HGNC Symbol;Acc:HGNC:6263]	0.81	0.00
863	ENSG00000116711	PLA2G4A	phospholipase A2 group IVA [Source:HGNC Symbol;Acc:HGNC:9035]	0.81	0.00
864	ENSG00000147174	ACRC	acidic repeat containing [Source:HGNC Symbol;Acc:HGNC:15805]	0.82	0.00
865	ENSG00000176076	KCNE5	potassium channel, voltage gated subfamily E regulatory beta subunit 5 [Source:HGNC Symbol;Acc:HGNC:6241]	0.82	0.00
866	ENSG00000145632	PLK2	polo-like kinase 2 [Source:HGNC Symbol;Acc:HGNC:19699]	0.82	0.00
867	ENSG00000102287	GABRE	gamma-aminobutyric acid (GABA) A receptor, epsilon [Source:HGNC Symbol;Acc:HGNC:4085]	0.82	0.00
868	ENSG00000198838	RYR3	ryanodine receptor 3 [Source:HGNC Symbol;Acc:HGNC:10485]	0.82	0.00
869	ENSG00000118515	SGK1	serum/glucocorticoid regulated kinase 1 [Source:HGNC Symbol;Acc:HGNC:10810]	0.83	0.00
870	ENSG00000125414	MYH2	myosin, heavy chain 2, skeletal muscle, adult [Source:HGNC Symbol;Acc:HGNC:7572]	0.83	0.00
871	ENSG00000173535	TNFRSF10C	tumor necrosis factor receptor superfamily member 10c [Source:HGNC Symbol;Acc:HGNC:11906]	0.83	0.00
872	ENSG00000104321	TRPA1	transient receptor potential cation channel, subfamily A, member 1 [Source:HGNC Symbol;Acc:HGNC:497]	0.83	0.00
873	ENSG00000186451	SPATA12	spermatogenesis associated 12 [Source:HGNC Symbol;Acc:HGNC:23221]	0.83	0.00
874	ENSG00000173404	INSM1	insulinoma associated 1 [Source:HGNC Symbol;Acc:HGNC:6090]	0.84	0.00
875	ENSG00000163053	SLC16A14	solute carrier family 16 member 14 [Source:HGNC Symbol;Acc:HGNC:26417]	0.85	0.00
876	ENSG00000130513	GDF15	growth differentiation factor 15 [Source:HGNC Symbol;Acc:HGNC:30142]	0.86	0.00
877	ENSG00000145642	FAM159B	family with sequence similarity 159 member B [Source:HGNC Symbol;Acc:HGNC:34236]	0.86	0.00
878	ENSG00000236581	STARD13-AS	STARD13 antisense RNA [Source:HGNC Symbol;Acc:HGNC:40873]	0.87	0.00
879	ENSG00000109255	NMU	neuromedin U [Source:HGNC Symbol;Acc:HGNC:7859]	0.87	0.00
880	ENSG00000174343	CHRNA9	cholinergic receptor, nicotinic alpha 9 [Source:HGNC Symbol;Acc:HGNC:14079]	0.88	0.00
881	ENSG00000118733	OLFM3	olfactomedin 3 [Source:HGNC Symbol;Acc:HGNC:17990]	0.88	0.00
882	ENSG00000104327	CALB1	calbindin 1 [Source:HGNC Symbol;Acc:HGNC:1434]	0.89	0.00
883	ENSG00000115758	ODC1	ornithine decarboxylase 1 [Source:HGNC Symbol;Acc:HGNC:8109]	0.89	0.00
884	ENSG00000243064	ABCC13	ATP binding cassette subfamily C member 13, pseudogene [Source:HGNC Symbol;Acc:HGNC:16022]	0.90	0.00
885	ENSG00000170961	HAS2	hyaluronan synthase 2 [Source:HGNC Symbol;Acc:HGNC:4819]	0.90	0.00
886	ENSG00000111452	ADGRD1	adhesion G protein-coupled receptor D1 [Source:HGNC Symbol;Acc:HGNC:19893]	0.91	0.00

887	ENSG00000014257	ACPP	acid phosphatase, prostate [Source:HGNC Symbol;Acc:HGNC:125]	0.92	0.00
888	ENSG00000162374	ELAVL4	ELAV like neuron-specific RNA binding protein 4 [Source:HGNC Symbol;Acc:HGNC:3315]	0.93	0.00
889	ENSG00000101977	MCF2	MCF.2 cell line derived transforming sequence [Source:HGNC Symbol;Acc:HGNC:6940]	0.95	0.00
890	ENSG00000130052	STARD8	StAR related lipid transfer domain containing 8 [Source:HGNC Symbol;Acc:HGNC:19161]	0.97	0.00
891	ENSG00000129538	RNASE1	ribonuclease, RNase A family, 1 (pancreatic) [Source:HGNC Symbol;Acc:HGNC:10044]	0.97	0.00
892	ENSG00000273036	FAM95C	family with sequence similarity 95 member C [Source:HGNC Symbol;Acc:HGNC:45272]	0.97	0.00
893	ENSG00000108691	CCL2	chemokine (C-C motif) ligand 2 [Source:HGNC Symbol;Acc:HGNC:10618]	0.98	0.00
894	ENSG00000138696	BMPR1B	bone morphogenetic protein receptor type IB [Source:HGNC Symbol;Acc:HGNC:1077]	0.98	0.00
895	ENSG00000187546	AGMO	alkylglycerol monooxygenase [Source:HGNC Symbol;Acc:HGNC:33784]	0.99	0.00
896	ENSG00000137948	BRDT	bromodomain, testis-specific [Source:HGNC Symbol;Acc:HGNC:1105]	1.02	0.00
897	ENSG00000168754	FAM178B	family with sequence similarity 178 member B [Source:HGNC Symbol;Acc:HGNC:28036]	1.03	0.00
898	ENSG00000157542	KCNJ6	potassium channel, inwardly rectifying subfamily J, member 6 [Source:HGNC Symbol;Acc:HGNC:6267]	1.05	0.00
899	ENSG00000143119	CD53	CD53 molecule [Source:HGNC Symbol;Acc:HGNC:1686]	1.05	0.00
900	ENSG00000232224	LINC00202-1	long intergenic non-protein coding RNA 202-1 [Source:HGNC Symbol;Acc:HGNC:24672]	1.06	0.00
901	ENSG00000076356	PLXNA2	plexin A2 [Source:HGNC Symbol;Acc:HGNC:9100]	1.07	0.00
902	ENSG00000138798	EGF	epidermal growth factor [Source:HGNC Symbol;Acc:HGNC:3229]	1.08	0.00
903	ENSG00000166292	TMEM100	transmembrane protein 100 [Source:HGNC Symbol;Acc:HGNC:25607]	1.09	0.00
904	ENSG00000280639	RP11-96C21.2		1.09	0.00
905	ENSG00000269964	MEI4	meiotic double-stranded break formation protein 4 [Source:HGNC Symbol;Acc:HGNC:43638]	1.10	0.00
906	ENSG00000100867	DHRS2	dehydrogenase/reductase (SDR family) member 2 [Source:HGNC Symbol;Acc:HGNC:18349]	1.10	0.00
907	ENSG00000178031	ADAMTSL1	ADAMTS like 1 [Source:HGNC Symbol;Acc:HGNC:14632]	1.13	0.00
908	ENSG00000111348	ARHGDIB	Rho GDP dissociation inhibitor (GDI) beta [Source:HGNC Symbol;Acc:HGNC:679]	1.13	0.00
909	ENSG00000162949	CAPN13	calpain 13 [Source:HGNC Symbol;Acc:HGNC:16663]	1.14	0.00
910	ENSG00000057657	PRDM1	PR domain containing 1, with ZNF domain [Source:HGNC Symbol;Acc:HGNC:9346]	1.15	0.00
911	ENSG00000243709	LEFTY1	left-right determination factor 1 [Source:HGNC Symbol;Acc:HGNC:6552]	1.16	0.00
912	ENSG00000165092	ALDH1A1	aldehyde dehydrogenase 1 family member A1 [Source:HGNC Symbol;Acc:HGNC:402]	1.17	0.00
913	ENSG00000150347	ARID5B	AT-rich interaction domain 5B [Source:HGNC Symbol;Acc:HGNC:17362]	1.17	0.00
914	ENSG00000225383	SFTA1P	surfactant associated 1, pseudogene [Source:HGNC Symbol;Acc:HGNC:18383]	1.19	0.01
915	ENSG00000143006	DMRTB1	DMRT like family B with proline rich C-terminal 1 [Source:HGNC Symbol;Acc:HGNC:13913]	1.20	0.00
916	ENSG00000182261	NLRP10	NLR family, pyrin domain containing 10 [Source:HGNC Symbol;Acc:HGNC:21464]	1.20	0.00
917	ENSG00000228592	D21S2088E	D21S2088E [Source:EntrezGene;Acc:266917]	1.22	0.01
918	ENSG00000168621	GDNF	glial cell derived neurotrophic factor [Source:HGNC Symbol;Acc:HGNC:4232]	1.22	0.00
919	ENSG00000163565	IFI16	interferon, gamma-inducible protein 16 [Source:HGNC Symbol;Acc:HGNC:5395]	1.23	0.00

920	ENSG00000243978	RGAG1	retrotransposon gag domain containing 1 [Source:HGNC Symbol;Acc:HGNC:29245]	1.25	0.00
921	ENSG00000198774	RASSF9	Ras association (RalGDS/AF-6) domain family (N-terminal) member 9 [Source:HGNC Symbol;Acc:HGNC:15739]	1.31	0.00
922	ENSG00000138449	SLC40A1	solute carrier family 40 (iron-regulated transporter), member 1 [Source:HGNC Symbol;Acc:HGNC:10909]	1.33	0.00
923	ENSG00000146047	HIST1H2BA	histone cluster 1, H2ba [Source:HGNC Symbol;Acc:HGNC:18730]	1.35	0.00
924	ENSG00000110195	FOLR1	folate receptor 1 (adult) [Source:HGNC Symbol;Acc:HGNC:3791]	1.39	0.00
925	ENSG00000125931	CITED1	Cbp/p300-interacting transactivator, with Glu/Asp rich carboxy-terminal domain, 1 [Source:HGNC Symbol;Acc:HGNC:1986]	1.47	0.00
926	ENSG00000231918	AC007682.1		1.49	0.00
927	ENSG00000205221	VIT	vitrin [Source:HGNC Symbol;Acc:HGNC:12697]	1.61	0.00
928	ENSG00000235268	KDM4E	lysine (K)-specific demethylase 4E [Source:HGNC Symbol;Acc:HGNC:37098]	1.93	0.00
929	ENSG00000156466	GDF6	growth differentiation factor 6 [Source:HGNC Symbol;Acc:HGNC:4221]	2.05	0.00
930	ENSG00000121858	TNFSF10	tumor necrosis factor superfamily member 10 [Source:HGNC Symbol;Acc:HGNC:11925]	2.09	0.00
931	ENSG00000164508	HIST1H2AA	histone cluster 1, H2aa [Source:HGNC Symbol;Acc:HGNC:18729]	2.09	0.00
932	ENSG00000105967	TFEC	transcription factor EC [Source:HGNC Symbol;Acc:HGNC:11754]	2.33	0.01
933	ENSG00000231944	PHKA1-AS1	PHKA1 antisense RNA 1 [Source:HGNC Symbol;Acc:HGNC:40446]	4.35	0.00

A.1.2. Gene ontology enrichment

Table (A.2) GO Component

	term id	term description	proteins	hits	p-value
1	GO:0005578	proteinaceous extracellular matrix	323	32	0.00000
2	GO:0031012	extracellular matrix	354	33	0.00000
3	GO:0044420	extracellular matrix component	116	16	0.00001
4	GO:0005588	collagen type V trimer	3	3	0.00006
5	GO:0005783	endoplasmic reticulum	1466	86	0.00012
6	GO:0005604	basement membrane	88	12	0.00017
7	GO:0043005	neuron projection	803	52	0.00031
8	GO:0005815	microtubule organizing center	557	39	0.00039
9	GO:0015630	microtubule cytoskeleton	938	58	0.00046
10	GO:0019867	outer membrane	174	17	0.00053
11	GO:0097458	neuron part	1026	62	0.00053
12	GO:0005741	mitochondrial outer membrane	146	15	0.00065
13	GO:0005813	centrosome	443	32	0.00075
14	GO:0044463	cell projection part	796	50	0.00079
15	GO:0036477	somatodendritic compartment	578	39	0.00079
16	GO:0043025	neuronal cell body	389	29	0.00080
17	GO:0031968	organelle outer membrane	168	16	0.00100
18	GO:0044297	cell body	436	31	0.00116
19	GO:0005788	endoplasmic reticulum lumen	189	17	0.00133
20	GO:0044429	mitochondrial part	826	50	0.00173
21	GO:0031233	intrinsic component of external side of plasma membrane	23	5	0.00174
22	GO:0044432	endoplasmic reticulum part	1014	59	0.00175

Table (A.3) GO Biological Process

	term id	term description	proteins	hits	p-value
1	GO:0022008	neurogenesis	1283	89	0.00000
2	GO:0048699	generation of neurons	1218	85	0.00000
3	GO:0030030	cell projection organization	910	69	0.00000
4	GO:0030198	extracellular matrix organization	341	36	0.00000
5	GO:0043062	extracellular structure organization	342	36	0.00000
6	GO:0044712	single-organism catabolic process	814	63	0.00000
7	GO:0048468	cell development	1470	96	0.00000
8	GO:0032989	cellular component morphogenesis	945	69	0.00000

9	GO:0000902	cell morphogenesis	869	65	0.00000
10	GO:2000026	regulation of multicellular organismal development	1400	90	0.00000
11	GO:0048858	cell projection morphogenesis	692	54	0.00000
12	GO:0031175	neuron projection development	617	49	0.00000
13	GO:0032990	cell part morphogenesis	711	54	0.00000
14	GO:0048646	anatomical structure formation involved in morphogenesis	914	64	0.00000
15	GO:0030182	neuron differentiation	901	63	0.00000
16	GO:0048666	neuron development	744	54	0.00000
17	GO:0006928	movement of cell or subcellular component	1211	77	0.00000
18	GO:0042325	regulation of phosphorylation	1162	74	0.00000
19	GO:0051094	positive regulation of developmental process	1018	67	0.00000
20	GO:0022603	regulation of anatomical structure morphogenesis	798	56	0.00000
21	GO:1902531	regulation of intracellular signal transduction	1267	79	0.00000
22	GO:0010647	positive regulation of cell communication	1439	87	0.00000
23	GO:0001932	regulation of protein phosphorylation	1083	70	0.00000
24	GO:0019220	regulation of phosphate metabolic process	1377	84	0.00000
25	GO:0040011	locomotion	1091	70	0.00001
26	GO:0051174	regulation of phosphorus metabolic process	1390	84	0.00001
27	GO:0009719	response to endogenous stimulus	1285	79	0.00001
28	GO:0044283	small molecule biosynthetic process	366	32	0.00001
29	GO:0048812	neuron projection morphogenesis	510	40	0.00001
30	GO:0051240	positive regulation of multicellular organismal process	1232	76	0.00001
31	GO:0000904	cell morphogenesis involved in differentiation	615	45	0.00001
32	GO:0035295	tube development	544	41	0.00001
33	GO:1901565	organonitrogen compound catabolic process	298	27	0.00002
34	GO:0045597	positive regulation of cell differentiation	746	51	0.00002

35	GO:0010243	response to organonitrogen compound	728	50	0.00002
36	GO:0031399	regulation of protein modification process	1448	84	0.00003
37	GO:0023056	positive regulation of signaling	1320	78	0.00003
38	GO:0019752	carboxylic acid metabolic process	736	50	0.00003
39	GO:0000003	reproduction	862	56	0.00003
40	GO:0006082	organic acid metabolic process	864	56	0.00003
41	GO:0016053	organic acid biosynthetic process	243	23	0.00004
42	GO:0046394	carboxylic acid biosynthetic process	243	23	0.00004
43	GO:0033554	cellular response to stress	1487	85	0.00004
44	GO:0043436	oxoacid metabolic process	848	55	0.00004
45	GO:0034405	response to fluid shear stress	36	8	0.00004
46	GO:0044703	multi-organism reproductive process	830	54	0.00004
47	GO:0071498	cellular response to fluid shear stress	19	6	0.00005
48	GO:0051241	negative regulation of multicellular organismal process	895	57	0.00005
49	GO:0043085	positive regulation of catalytic activity	1302	76	0.00005
50	GO:0032504	multicellular organism reproduction	719	48	0.00006
51	GO:0044711	single-organism biosynthetic process	1206	71	0.00007
52	GO:1901575	organic substance catabolic process	1385	79	0.00008
53	GO:1901700	response to oxygen-containing compound	1214	71	0.00009
54	GO:0009967	positive regulation of signal transduction	1194	70	0.00009
55	GO:0071495	cellular response to endogenous stimulus	938	58	0.00009
56	GO:0070555	response to interleukin-1	75	11	0.00009
57	GO:0010212	response to ionizing radiation	130	15	0.00009
58	GO:0061564	axon development	477	35	0.00010
59	GO:0007409	axonogenesis	459	34	0.00011
60	GO:0071417	cellular response to organonitrogen compound	459	34	0.00011
61	GO:0048667	cell morphogenesis involved in neuron differentiation	498	36	0.00011
62	GO:2000425	regulation of apoptotic cell clearance	8	4	0.00012
63	GO:0006942	regulation of striated muscle contraction	90	12	0.00012
64	GO:1901698	response to nitrogen compound	800	51	0.00012

65	GO:0044702	single organism reproductive process	970	59	0.00013
66	GO:0044248	cellular catabolic process	1317	75	0.00013
67	GO:0016054	organic acid catabolic process	180	18	0.00013
68	GO:0046395	carboxylic acid catabolic process	180	18	0.00013
69	GO:0086004	regulation of cardiac muscle cell contraction	42	8	0.00013
70	GO:0048609	multicellular organismal reproductive process	701	46	0.00013
71	GO:2000241	regulation of reproductive process	92	12	0.00015
72	GO:0019953	sexual reproduction	684	45	0.00015
73	GO:0030334	regulation of cell migration	564	39	0.00015
74	GO:0022414	reproductive process	1083	64	0.00015
75	GO:0034616	response to laminar fluid shear stress	15	5	0.00015
76	GO:0055117	regulation of cardiac muscle contraction	79	11	0.00015
77	GO:0009887	organ morphogenesis	809	51	0.00015
78	GO:0051345	positive regulation of hydrolase activity	769	49	0.00016
79	GO:0055114	oxidation-reduction process	917	56	0.00017
80	GO:0051336	regulation of hydrolase activity	1154	67	0.00017
81	GO:0051270	regulation of cellular component movement	669	44	0.00017
82	GO:1901564	organonitrogen compound metabolic process	1465	81	0.00018
83	GO:0060341	regulation of cellular localization	1092	64	0.00019
84	GO:0031667	response to nutrient levels	397	30	0.00019
85	GO:0060322	head development	652	43	0.00019
86	GO:0043547	positive regulation of GTPase activity	455	33	0.00020
87	GO:0051130	positive regulation of cellular component organization	1057	62	0.00023
88	GO:0045595	regulation of cell differentiation	1343	75	0.00023
89	GO:1901699	cellular response to nitrogen compound	500	35	0.00025
90	GO:1903115	regulation of actin filament-based movement	46	8	0.00025
91	GO:0009991	response to extracellular stimulus	425	31	0.00028
92	GO:0009063	cellular amino acid catabolic process	99	12	0.00029
93	GO:0008016	regulation of heart contraction	160	16	0.00030
94	GO:0007283	spermatogenesis	466	33	0.00030

95	GO:0033280	response to vitamin D	26	6	0.00031
96	GO:0048232	male gamete generation	467	33	0.00032
97	GO:0040012	regulation of locomotion	647	42	0.00032
98	GO:0048585	negative regulation of response to stimulus	1202	68	0.00032
99	GO:0071305	cellular response to vitamin D	10	4	0.00033
100	GO:1902547	regulation of cellular response to vascular endothelial growth factor stimulus	10	4	0.00033
101	GO:1901652	response to peptide	432	31	0.00036
102	GO:0051046	regulation of secretion	612	40	0.00038
103	GO:0009083	branched-chain amino acid catabolic process	18	5	0.00039
104	GO:0010563	negative regulation of phosphorus metabolic process	473	33	0.00040
105	GO:0045936	negative regulation of phosphate metabolic process	473	33	0.00040
106	GO:0007584	response to nutrient	164	16	0.00040
107	GO:0097305	response to alcohol	303	24	0.00041
108	GO:0070848	response to growth factor	614	40	0.00041
109	GO:0071363	cellular response to growth factor stimulus	595	39	0.00043
110	GO:0071377	cellular response to glucagon stimulus	38	7	0.00043
111	GO:0031960	response to corticosteroid	149	15	0.00043
112	GO:0048514	blood vessel morphogenesis	361	27	0.00045
113	GO:2000427	positive regulation of apoptotic cell clearance	5	3	0.00048
114	GO:2000145	regulation of cell motility	599	39	0.00049
115	GO:0044057	regulation of system process	401	29	0.00049
116	GO:0001822	kidney development	236	20	0.00052
117	GO:1903530	regulation of secretion by cell	561	37	0.00053
118	GO:0043087	regulation of GTPase activity	501	34	0.00053
119	GO:0007417	central nervous system development	792	48	0.00058
120	GO:1901342	regulation of vasculature development	204	18	0.00060
121	GO:0044282	small molecule catabolic process	240	20	0.00064
122	GO:0060429	epithelium development	903	53	0.00064
123	GO:0032963	collagen metabolic process	79	10	0.00065
124	GO:0006935	chemotaxis	588	38	0.00066

125	GO:0042330	taxis	588	38	0.00066
126	GO:0051960	regulation of nervous system development	630	40	0.00068
127	GO:0003012	muscle system process	242	20	0.00071
128	GO:0007411	axon guidance	373	27	0.00075
129	GO:0097485	neuron projection guidance	373	27	0.00075
130	GO:0051384	response to glucocorticoid	141	14	0.00075
131	GO:0007276	gamete generation	572	37	0.00076
132	GO:0009081	branched-chain amino acid metabolic process	21	5	0.00085
133	GO:0044782	cilium organization	176	16	0.00086
134	GO:0022617	extracellular matrix disassembly	112	12	0.00090
135	GO:0010811	positive regulation of cell-substrate adhesion	97	11	0.00091
136	GO:0042326	negative regulation of phosphorylation	359	26	0.00092
137	GO:0032226	positive regulation of synaptic transmission, dopaminergic	6	3	0.00093
138	GO:0060385	axonogenesis involved in innervation	6	3	0.00093
139	GO:0072001	renal system development	248	20	0.00096
140	GO:0010810	regulation of cell-substrate adhesion	161	15	0.00096
141	GO:0044087	regulation of cellular component biogenesis	601	38	0.00098
142	GO:0043434	response to peptide hormone	401	28	0.00104
143	GO:0034612	response to tumor necrosis factor	114	12	0.00106
144	GO:0044259	multicellular organismal macromolecule metabolic process	84	10	0.00106
145	GO:0030574	collagen catabolic process	70	9	0.00107
146	GO:0001933	negative regulation of protein phosphorylation	326	24	0.00114
147	GO:0051248	negative regulation of protein metabolic process	949	54	0.00116
148	GO:0072330	monocarboxylic acid biosynthetic process	164	15	0.00116
149	GO:0007420	brain development	609	38	0.00125
150	GO:0010927	cellular component assembly involved in morphogenesis	236	19	0.00130
151	GO:0050804	modulation of synaptic transmission	236	19	0.00130
152	GO:0031669	cellular response to nutrient levels	183	16	0.00131
153	GO:0042594	response to starvation	183	16	0.00131

154	GO:0009725	response to hormone	802	47	0.00132
155	GO:0035513	oxidative RNA demethylation	2	2	0.00137
156	GO:0035552	oxidative single-stranded DNA demethylation	2	2	0.00137
157	GO:0035553	oxidative single-stranded RNA demethylation	2	2	0.00137
158	GO:0042822	pyridoxal phosphate metabolic process	2	2	0.00137
159	GO:1900623	regulation of monocyte aggregation	2	2	0.00137
160	GO:1900625	positive regulation of monocyte aggregation	2	2	0.00137
161	GO:1901074	regulation of engulfment of apoptotic cell	2	2	0.00137
162	GO:2000243	positive regulation of reproductive process	34	6	0.00140
163	GO:0014070	response to organic cyclic compound	740	44	0.00141
164	GO:0000724	double-strand break repair via homologous recombination	73	9	0.00145
165	GO:0090257	regulation of muscle system process	185	16	0.00146

Table (A.4) GO Molecular Function

	term id	term description	proteins	hits	p-value
1	GO:0016491	oxidoreductase activity	638	44	0.00000
2	GO:0032559	adenyl ribonucleotide binding	1450	79	0.00000
3	GO:0030554	adenyl nucleotide binding	1456	79	0.00000
4	GO:0030234	enzyme regulator activity	841	51	0.00001
5	GO:0005524	ATP binding	1415	75	0.00001
6	GO:0005539	glycosaminoglycan binding	190	19	0.00001
7	GO:0050839	cell adhesion molecule binding	128	15	0.00002
8	GO:0098772	molecular function regulator	1094	61	0.00002
9	GO:0004222	metalloendopeptidase activity	104	13	0.00003
10	GO:0035514	DNA demethylase activity	3	3	0.00003
11	GO:0043734	DNA-N1-methyladenine dioxygenase activity	3	3	0.00003
12	GO:0008047	enzyme activator activity	423	30	0.00005
13	GO:0005509	calcium ion binding	663	41	0.00005
14	GO:0050660	flavin adenine dinucleotide binding	70	10	0.00008
15	GO:0030695	GTPase regulator activity	274	22	0.00008
16	GO:0060589	nucleoside-triphosphatase regulator activity	296	23	0.00009

17	GO:0005178	integrin binding	74	10	0.00012
18	GO:0005096	GTPase activator activity	245	20	0.00014
19	GO:0008201	heparin binding	144	14	0.00023
20	GO:0005102	receptor binding	1038	54	0.00034
21	GO:0005518	collagen binding	43	7	0.00040
22	GO:0005215	transporter activity	1055	54	0.00050
23	GO:0016462	pyrophosphatase activity	715	40	0.00051
24	GO:0016818	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	718	40	0.00055
25	GO:0016817	hydrolase activity, acting on acid anhydrides	719	40	0.00056
26	GO:0008762	UDP-N-acetylmuramate dehydrogenase activity	6	3	0.00061
27	GO:0017111	nucleoside-triphosphatase activity	679	38	0.00069
28	GO:0016772	transferase activity, transferring phosphorus-containing groups	849	45	0.00072
29	GO:0016301	kinase activity	705	39	0.00073
30	GO:0001071	nucleic acid binding transcription factor activity	973	50	0.00074
31	GO:0003700	transcription factor activity, sequence-specific DNA binding	973	50	0.00074
32	GO:0022892	substrate-specific transporter activity	902	47	0.00080
33	GO:0008753	NADPH dehydrogenase (quinone) activity	2	2	0.00103
34	GO:0035515	oxidative RNA demethylase activity	2	2	0.00103
35	GO:0051747	cytosine C-5 DNA demethylase activity	2	2	0.00103
36	GO:0097363	protein O-GlcNAc transferase activity	2	2	0.00103
37	GO:0016788	hydrolase activity, acting on ester bonds	625	35	0.00109
38	GO:0051213	dioxygenase activity	82	9	0.00125
39	GO:0031267	small GTPase binding	154	13	0.00143
40	GO:0048037	cofactor binding	253	18	0.00144
41	GO:0005525	GTP binding	362	23	0.00157
42	GO:0016773	phosphotransferase activity, alcohol group as acceptor	640	35	0.00163
43	GO:0003840	gamma-glutamyltransferase activity	8	3	0.00164
44	GO:0008289	lipid binding	594	33	0.00170
45	GO:1901681	sulfur compound binding	199	15	0.00199

46	GO:0016887	ATPase activity	347	22	0.00201

A.1.3. R code for RNAseq analysis

RNAseq analysis

Required packages: Bioconductor, edgeR, STRINGdb, biomaRT, pylr

Differential gene expression with edgeR

Making the "wide"-object that can be used for DEG with EdgeR:

```
setwd("D:/Dropbox/Tanja's PhD/lab stuff/RNAseq ChiPS4 GNS/RNAseq analysis/Analysis Tanja")
full<-read.csv("fullResults.csv", header=T, sep=";")


colnames(full)

fullu<-unique(full[, -c(1, which(colnames(full)=="entrezgene"))])

wide<-fullu[,5:16]
#View(wide)
rownames(wide)<-fullu$ensembl_gene_id
#View(wide)
```

These commands are used to run the differential gene expression analysis:

```
library(edgeR)

dge <- DGEList(wide,group=as.factor(substr(colnames(wide),1,2))
+ )
design <- model.matrix(~dge$samples$group -1)
colnames(design) <- levels(dge$samples$group)
dge <- calcNormFactors(dge)
dge <- estimateDisp(dge,design)
fit <- glmFit(dge,design)
lrt <- glmLRT(fit,contrast=c(-1,1))
ttn<-topTags(lrt,100000)$table
```

StringDB & biomaRT

This frist chunk generates the environment:

StringDB analysis:

```
#biocLite("STRINGdb")
require(STRINGdb)
string_db <- STRINGdb$new( version="10", species=9606) #Builds the environment

rl<-read.csv("sorted_hits_Stringdb_TD.csv", header=T, sep=",", stringsAsFactors = FALSE)

rl<-unique(rl)
rl<-na.omit(rl)
string_db <- STRINGdb$new( version="10", species=9606)
```

```
rl_mapped <- string_db$map(rl, "entrezgene", removeUnmappedRows = TRUE )

rl_mapped_sig<-string_db$add_diff_exp_color(rl_mapped, logFcColStr = "logFC")

#to only get proteins with FDR<0.05:
rl_mapped_sig<-string_db$add_diff_exp_color(subset(rl_mapped, FDR<0.05,
                                                   logFcColStr = "logFC")

#in the end I used these settings: abs(logFC)>0.85
rl_mapped_sig<-string_db$add_diff_exp_color(subset (rl_mapped, abs(logFC)>0.85),
                                            logFcColStr = "logFC")

payload_id <- string_db$post_payload(rl_mapped_sig$STRING_id,  colors=rl_mapped_sig$color)
string_db$plot_network(rl_mapped_sig$STRING_id[1:200], #show 200 proteins
                       payload_id=payload_id )
```

Gene ontology analysis

The code below allows to retrieve the genes within a GO hit that are in our DEG list. Alternatively, the web version of GO can be used.

```
rl <- read.csv("topAnnotatedhitsR.csv", header=T, sep=";", stringsAsFactors = FALSE)


goGenes <- getBM(attributes=c("ensembl_gene_id","external_gene_name"),
             filters="go_parent_term",
             values="GO:0048699", #put the desired GO term id here
             mart=ens)


goGenes<-getBM(attributes=c("ensembl_gene_id","external_gene_name"),
               filters=c("go_parent_term","go_evidence_code"),
               values=list("GO:0007417","IEA"), mart=ens)



hitsProcess<-merge(goGenes, rl, by="ensembl_gene_id")
write.xlsx(hitsProcess, "hits in GO terms/GO0007417.xlsx")
```

A.2. Publications

STEM CELLS®

EMBRYONIC STEM CELLS/INDUCED PLURIPOTENT STEM CELLS

Elevated *O*-GlcNAc Levels Activate Epigenetically Repressed Genes and Delay Mouse ESC Differentiation Without Affecting Naïve to Primed Cell Transition

CHRISTOPHER M. SPEAKMAN,[a] TANJA C.E. DOMKE,[a] WIKROM WONGPAIBOONWATTANA,[a] KELLY SANDERS,[b] MANIKHANDAN MUDALIAR,[c,d] DAAN M.F. VAN AALTEN,[b] GEOFFREY J. BARTON,[c] MARIOS P. STAVRIDIS[a]

Key Words. Embryonic stem cells • Cell differentiation • *O*-GlcNAc • Post-translational protein modification • Signal transduction • Oligonucleotide microarrays

[a]Division of Cancer Research, Medical Research Institute, College of Medicine, Dentistry and Nursing; [b]Division of Molecular Microbiology, College of Life Sciences, and [c]Division of Computational Biology, College of Life Sciences, University of Dundee, Dundee, United Kingdom; [d]Glasgow Polyomics, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom

Correspondence: Marios P. Stavridis, Ph.D., Medical Research Institute, Division of Cancer Research, Jacqui Wood Cancer Centre, University of Dundee, James Arrott Drive, Ninewells Hospital & Medical School, Dundee DD1 9SY, U.K. Telephone: 44-1382-383388; Fax: 44-1382-386419; e-mail: m.stavridis@dundee.ac.uk

Received July 17, 2013; accepted for publication May 10, 2014; first published online in STEM CELLS *EXPRESS* June 4, 2014.

1066-5099/2014/$30.00/0

http://dx.doi.org/10.1002/stem.1761

ABSTRACT

The differentiation of mouse embryonic stem cells (ESCs) is controlled by the interaction of multiple signaling pathways, typically mediated by post-translational protein modifications. The addition of *O*-linked *N*-acetylglucosamine (*O*-GlcNAc) to serine and threonine residues of nuclear and cytoplasmic proteins is one such modification (*O*-GlcNAcylation), whose function in ESCs is only now beginning to be elucidated. Here, we demonstrate that the specific inhibition of *O*-GlcNAc hydrolase (Oga) causes increased levels of protein *O*-GlcNAcylation and impairs differentiation of mouse ESCs both in serum-free monolayer and in embryoid bodies (EBs). Use of reporter cell lines demonstrates that Oga inhibition leads to a reduction in the number of Sox1-expressing neural progenitors generated following induction of neural differentiation as well as maintained expression of the ESC marker Oct4 (Pou5f1). In EBs, expression of mesodermal and endodermal markers is also delayed. However, the transition of naïve cells to primed pluripotency indicated by Rex1 (Zfp42), Nanog, Esrrb, and Dppa3 downregulation and Fgf5 upregulation remains unchanged. Finally, we demonstrate that increased *O*-GlcNAcylation results in upregulation of genes normally epigenetically silenced in ESCs, supporting the emerging role for this protein modification in the regulation of histone modifications and DNA methylation. STEM CELLS *2014;32:2605–2615*

INTRODUCTION

Over the last decade there have been major advances in our understanding of the mechanisms controlling embryonic stem cell (ESC) behavior in response to the changing extracellular environment [1]. These mechanisms largely involve engagement of signal transduction relays that operate by post-translational modifications of proteins. Under standard conditions, ESCs grow as a mixture of "naïve" and "primed" cells. Signaling mediated by Erk1/2 target phosphorylation has recently been implicated in regulating the transition between these two states and initiation of differentiation [2]. Reversible protein modification by addition of *O*-linked *N*-acetylglucosamine to serine or threonine residues (*O*-GlcNAcylation) was first described 30 years ago [3] and occurs with similar time scales, dynamics, and stoichiometry as protein phosphorylation, with which it sometimes competes. *O*-GlcNAcylation is found in all higher eukaryotes tested to date and has been implicated in development, epigenetic regulation, and diseases such as diabetes and Alzheimer's [4]. Its addition and removal are catalyzed by one transferase (Ogt) and one hydrolase (Oga; also known as Mgea5), respectively.

To date there have been very few studies on *O*-GlcNAc function in ESCs, although accumulating evidence suggests a critical role for this modification. Elimination of Ogt in mouse ESCs or conditional deletion in somatic cells leads to death, consistent with an essential role in all cell types [5, 6]. Ogt has recently been identified as a protein partner of the essential ESC transcription factor Oct4 (Pou5f1) by three independent studies [7–9]. Furthermore, Oct4 has been shown to be modified by *O*-GlcNAc [10], and this was demonstrated to be important for regulation of a subset of its targets in ESCs and during reprogramming [11]. A previous study has suggested that increased *O*-GlcNAcylation in ESCs prevents differentiation along the cardiac lineage in spontaneously differentiating embryoid bodies (EBs) [12]; however, the mechanism for this or the stage at

which differentiation stalled was not determined. Ogt is a mammalian homolog of the *Drosophila super sex combs* (scx) gene, a member of the Polycomb group of transcriptional repressors [13, 14] and has established roles in gene repression [15]. Furthermore, recent studies implicate *O*-GlcNAc transferase in the regulation of the Tet epigenetic modifiers [16–18], suggesting that the *O*-GlcNAc modification and the enzymes controlling it regulate chromatin in pluripotent cells by multiple mechanisms.

In this study, we investigate in detail the effects of increased *O*-GlcNAcylation on neural differentiation of mouse ESCs. Our results show that Oga inhibition leads to delayed onset of differentiation although the transition of naïve cells to primed pluripotency proceeds unhindered. We also present a genome-wide gene expression analysis of ESCs and differentiating cells treated with a highly specific Oga inhibitor, revealing that increased *O*-GlcNAc levels control the expression of distinct groups of genes in ESCs associated with a recently described subpopulation resembling the two-cell-stage embryo. Upregulation of these gene sets is consistent with a disruption of the normal transcriptional repression programme operating in pluripotent cells.

Materials and Methods

ESC Culture and Differentiation

Mouse ESCs were cultured without feeders on 0.1% gelatin-coated plastics in Glasgow Minimal Essential Medium with 10% serum and Leukemia Inhibitory Factor (LIF) [19] and differentiation was performed according to our previous protocol [20]. The cell lines used were derivatives of E14Tg2aIV (46C [21], OCRG9 [22]) or Oct4GiP [23]. GlcNAcstatin C (GNS) was obtained from GlycoBioChem (Dundee, UK; www.glycobiochem.com) and used at 1 μM unless otherwise specified. For flow cytometry, cells were dissociated using Accutase (Millipore, Billerica, MA; www.emdmillipore.com) and resuspended in PBS/1% bovine serum albumin (BSA) containing 1 μg/ml propidium iodide (PI; Sigma, St. Louis, MO; www.sigmaaldrich.com). Cellular debris and PI-positive cells were excluded from analysis. For clonal analysis, 600/100 cells were plated per 10 cm dish/well of six-well plate and cultured for 6–8 days, then fixed, and stained using an Alkaline phosphatase staining kit (Sigma).

siRNA Induced Knockdown of Oga and Ogt

Cells were seeded into six-well dishes (2×10^5 cells/well). While still in suspension, cells were transfected with SmartPool siRNAs targeting Oga or Ogt or with a nontargeting siRNA pool (50 pmol per well, Thermo Scientific, Waltham, MA; www.thermoscientific.com) using Lipofectamine RNAiMAX (Life Technologies, Carlsbad, CA; www.lifetechnologies.com) (5 μl per well). Cells were transfected again after 24 hours.

Western Blotting

Cells were cultured or differentiated in 9 cm dishes and treatments were for 24 hours unless otherwise specified. Lysis was performed on ice in whole-cell lysis buffer (150 mM sodium chloride, 1.0% Nonidet P40, 50 mM Tris, pH 8.0 with 1 μM GNS, Complete protease inhibitor, and PhosStop phosphatase inhibitor cocktail tablets from Roche, Basel, Switzerland; www.roche.com). For nuclear/cytoplasmic fractionation, cells were lysed on ice in Buffer A (10 mM HEPES, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM Dithiothreitol (DTT), 0.05% Nonidet P40 pH 7.9) for 10 minutes and centrifuged for 10 minutes at 3,000 rpm in a microfuge at 4°C. The supernatant was removed as a cytoplasmic fraction and the pellet was resuspended in Buffer B (5 mM HEPES, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.5 mM DTT, 26% glycerol [vol/vol], pH 7.9) by micropestle homogenization, NaCl added to 300 mM, and incubated on ice for 30 minutes. A final spin at 12,000*g* was performed to separate insoluble material from the nuclear fraction. Protein concentration was determined by Coomassie protein assay (Thermo Scientific) and 10–30 μg of protein was loaded per lane of Life Technologies NuPage gels and transferred to Polyvinylidene fluoride membrane (Millipore). Membranes were blocked in either 5% Milk (Marvel, Premier Foods, St. Albans, UK; www.premierfoods.co.uk) in TBST or 1%–5% BSA (Millipore) in TBST (all other chemicals from Sigma). Antibodies were incubated in blocking buffer overnight: anti-*O*-GlcNAc, CTD110.6 (Covance, Princeton, NJ; www.covance.com, 1/5,000), or RL-2 (Santa Cruz Biotechnology, Dallas, TX; www.scbt.com, 1/1,000); beta actin (Abcam, Cambridge, UK; www.abcam.co.uk, 1/2,000); phosphoErk1/2 (Cell Signaling Technologies, Beverly, MA; www.cellsignal.com, 1/1,000); Gsk3α and phosphoGsk3α/β (Cell Signaling, 1/1,000); phospho-Serine and phospho-Threonine (Cell Signaling, 1/1,000); Oct4 (Abcam 1/1,000); Sox2 (Abcam 1/1,000). For the phospho-kinase antibody array (R&D Systems, Minneapolis, MN; www.rndsystems.com) 500 μg of total cell lysate was used according to manufacturers instructions and the arrays were developed using LumiGLO from Cell Signaling and imaged on a Fuji LAS3000 mini scanner. Densitometry was performed with ImageJ.

Immunoprecipitation

Whole cell lysate (500 μg) was incubated overnight with a mixture of succinylated wheat germ agglutinin-agarose and protein A sepharose beads in the presence of antibody RL-2 (anti-*O*-GlcNAc). After washing the bound proteins were eluted in 25 μl of 2× LDS loading buffer (Life Technologies), boiled for 5 minutes, and analyzed by Western blotting.

Immunocytochemistry

Cells were fixed in 4% paraformaldehyde (PFA) for 15 minutes, permeabilized in PBS, 0.1% Tween (PBST), and incubated in PBST +5% BSA (blocking buffer). Primary antibody (goat anti-Oct4, Santa Cruz) was diluted in blocking buffer at 1/50 and incubated overnight at 4°C. Subsequently, cells were incubated in fluorescent secondary antibody (AlexaFluor conjugate, Life Technologies, 1/1,000) and 300 nM 4′,6-diamidino-2-phenylindole for 1 hour.

Two-Dimensional-PAGE

Nuclear samples extracted from ESCs treated with LIF, or 1–4 days treatment with N2B27, were enzymatically labeled using Click-IT *O*-GlcNAc labeling system (Life Technologies) according to the manufacturers instructions. Samples were chloroform-methanol precipitated and resuspended in 7 M urea, 2 M thiourea, 4% CHAPS, 1% ASB-14, and 0.5% ampholytes. Passive rehydration was performed overnight followed by isoelectric

focusing using pH 3–10 NL IPG strips for 4,200 vh. Preceding equilibration, samples were separated by SDS-PAGE using 4%–12% bis/tris gels and transferred to Polyvinylidene fluoride for detection by streptavidin-HRP. Ten prominent spots were excised from duplicate gels and analyzed by MALDI-TOF mass spectrometry.

Quantitative RT-PCR

RNA was extracted from cells grown in six-well plates using Nucleospin II RNA kit (Macherey-Nagel, Dueren, Germany; www.mn-net.com). DNase treatment was performed on-column during RNA extraction. One microgram of RNA was reverse transcribed with the qScript cDNA synthesis kit (Quanta Biosciences, Gaithersburg, MD; www.quantabio.com). PCR was performed on a BioRad iCycler or an AB 7500 using PerfeCTa SYBR Green FastMix for iQ or PerfeCTa SYBR Green FastMix Low Rox (both Quanta Biosciences), respectively. Primer sequences are shown in Supporting Information Table S1. Relative quantitation was performed using the method by Pfaffl [24] and β-actin as a reference gene from technical duplicates or triplicates of at least three independent experiments.

Microarray Analysis

RNA samples (from four independent experiments) were processed and hybridized to Affymetrix Gene 1.1ST arrays by Ark Genomics (www.ark-genomics.org) and the raw array expression data were obtained as CEL files. For quality control and probe sets annotations, the annotations files (Release 32, dated 23-06-2011) downloaded from Affymetrix (Santa Clara, CA; www.affymetrix.com) were used. Background noise control (Detected Above Background), Robust Multi-array Average normalization, and summarization of probe set level data into transcript clusters were carried out using Affymetrix Power Tools. Quality analysis and differential expression analyses were performed in Partek GS 6.5 (version 6.11.0321) software and R (version 2.13.1)—Bioconductor using Limma [25], and RankProd [26] packages [27]. The microarray data have been deposited in EBI ArrayExpress under accession number E-MEXP-3593. For gene set enrichment analysis (GSEA) GSEA v2.0.13 was used, along with gene sets from MSigDB, GenesigDB, or custom made ones from the literature (Supporting Information Table S5). Genes were ranked by the Signal2Noise metric and the weighted2 enrichment statistic was used over 1,000 gene set permutations. A false discovery rate q-value cut-off of 0.05 was applied.

Results

Protein O-GlcNAcylation Delays Mouse ESC Differentiation

We decided to investigate the changes in *O*-GlcNAc signaling during ESC neural differentiation. Global *O*-GlcNAc levels decline slightly during the first few days of differentiation, as measured by Western blotting with the *O*-GlcNAc-specific antibody CTD110.6 (Fig. 1A). However, this method may miss changes in the *O*-GlcNAcylation of lower abundance proteins. To test for more subtle changes in protein *O*-GlcNAcylation, we performed a two-dimensional-PAGE-Western blot analysis of nuclear extracts of ESCs and cells during the first 4 days of monolayer differentiation using chemoenzymatic labeling of *O*-GlcNAcylated proteins (Fig. 1B and Supporting Information Fig. S1A). The results revealed a dynamic pattern of nuclear *O*-GlcNAcylated proteins, with a dramatic change during the first 24 hours of differentiation. As the *O*-GlcNAcylation on the majority of the nuclear proteins disappeared following 24 hours of differentiation, we hypothesized that removal of the *O*-GlcNAc modification from some ESC protein(s) may be required for differentiation onset. We also picked 10 spots showing dramatic regulation from the Day 1 gel and identified them by mass spectrometry. Our analysis identified 26 unique proteins. Although we were not able to positively identify *O*-GlcNAc modification on any of these proteins due to the instrumentation used, we note that 13 of the 26 had previously been identified as *O*-GlcNAc modified by other screens (Supporting Information Table S2).

We then sought to determine whether the *O*-GlcNAc levels in ESCs can be modulated by inhibiting the Oga enzyme activity that removes *O*-GlcNAc using GNS, a highly potent and specific small molecule Oga inhibitor that exhibits 164-fold selectivity for Oga when tested against the closely related lysosomal hexosaminidases HexA/B [28] and a competitive inhibitor of the *O*-GlcNAc transferase Ogt (4Ac5SGlcNAc) [29]. Treatment of ESCs with GNS resulted in a dose-dependent increase in cellular *O*-GlcNAc levels (Fig. 1C) including *O*-GlcNAcylation of the transcription factor Sox2 on S248 (Fig. 1D), with no discernable effects on cell viability or proliferation even after prolonged treatment (Fig. 1E, 1F). Similar results were obtained by use of siRNA (Supporting Information Fig. S1B). Inhibition of Ogt conversely led to cell death within 4–5 days in culture (Supporting Information Fig. S1C), consistent with the previous observation that this protein is essential for ESC viability [5] (Fig. 3A).

We then treated murine 46C ESCs (engineered to express the green fluorescent protein [GFP] from the endogenous neural-specific *Sox1* locus [21]) with GNS during monolayer differentiation into neural cells [20]. Flow cytometry showed that a significantly smaller proportion of neural progenitors were generated during ESC differentiation in the presence of GNS (24.6% in GNS vs. 32.2% in DMSO at day 3, a reduction of 23.6%; $n = 4$, $p = .0118$) (Fig. 2A, Supporting Information Videos S1 and S2). Similar effects were seen using a structurally distinct Oga inhibitor, Thiamet G (21.4% reduction at day 3). *Sox1* mRNA levels were similarly reduced (Supporting Information Fig. S2A). This decrease in neural progenitors could reflect a bias of differentiation against neural fate, or be due to a more general effect on the onset of differentiation irrespective of lineage. To test this differentiating 46C cells were assessed for expression of the ESC marker Oct4. GNS-treated ESCs remained Oct4 positive when vehicle-treated control cells have largely lost Oct4 immunoreactivity and instead expressed Sox1GFP (Fig. 2B). We obtained the same result using the transgenic *Oct4*-GFP reporter ESC line Oct4GiP (Fig. 2C), suggesting a delay in differentiation onset in the presence of GNS. Consistently, we also observed a decrease in endoderm (BMP2) and mesoderm (Brachyury, Eomesodermin) markers following EB differentiation (Fig. 2D–2F). Interestingly, primitive endoderm differentiation measured by RT-qPCR for Sox17 and Gata6 appears to be unaffected by GNS (Supporting Information Fig. S2B, S2C).

ESC self-renewal can be maintained through the action of the transcription factor Stat3, operating downstream of the

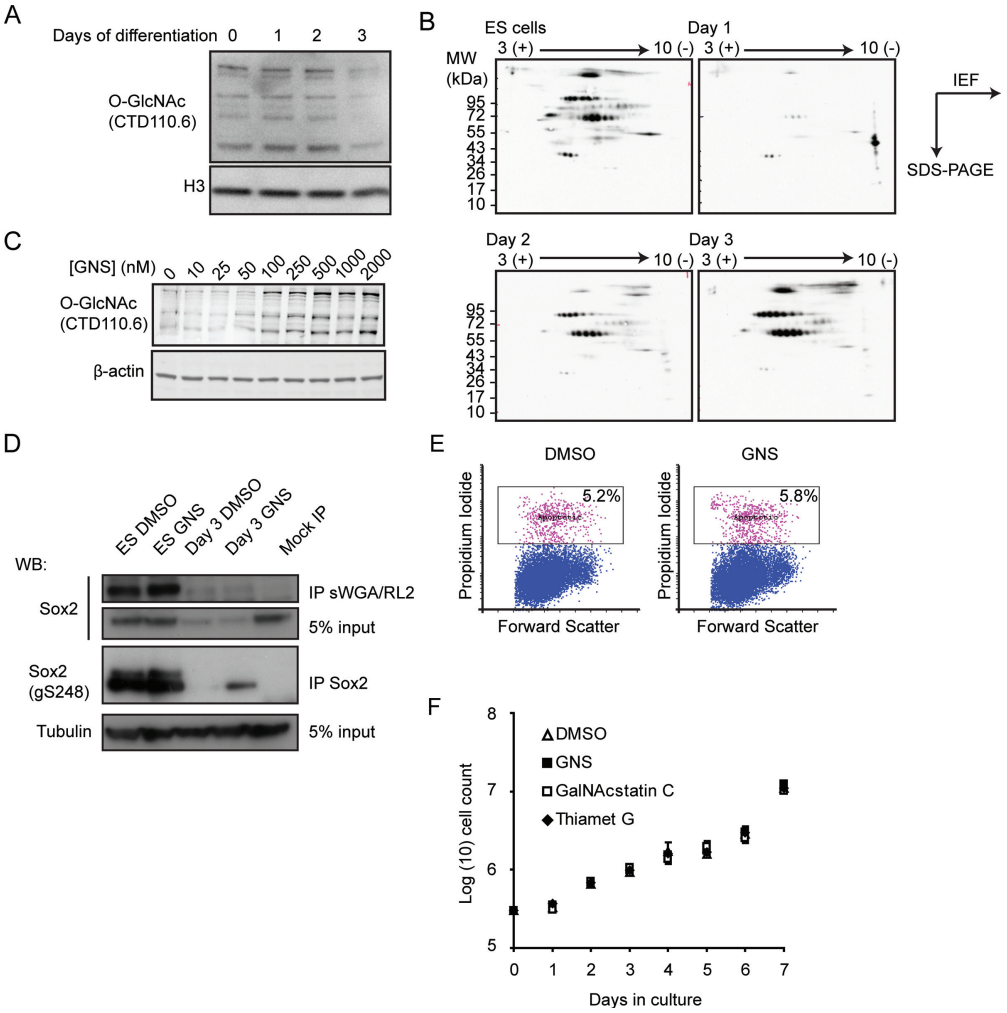

Figure 1. Global protein *O*-GlcNAcylation in embryonic stem cells (ESCs). **(A):** Whole-cell lysates of ESCs and cells at various stages of neural differentiation were blotted for *O*-GlcNAc levels. No major changes are detectable in the most abundant bands. **(B):** Two-dimensional gels of nuclear extracts blotted for *O*-GlcNAcylation. The overall level of *O*-GlcNAc is reduced in differentiation conditions whereas a cluster of basic proteins is prominent in the Day 1 sample. **(C):** Dose-dependent increase in protein *O*-GlcNAcylation after 24 hours treatment with 1 μM GNS. **(D):** Undifferentiated and 3-day neural differentiated ESCs in DMSO or GNS were immunoprecipitated with antibody against *O*-GlcNAc (RL-2) and succinylated WGA beads or antibody against Sox2 then blotted for Sox2 or Sox2 GlcNAcylated at S248. GNS increases *O*-GlcNAc levels on Sox2, especially on day 3. **(E):** Measurement of apoptotic cells (pink) after 24 hours of vehicle (DMSO) or 1 μM GNS treatment shows no difference in apoptosis levels. **(F):** ESC proliferation in vehicle (DMSO), the Oga inhibitors Thiamet G, and GNS or its inactive stereoisomer, GalNacstatin C (all at 1 μM). Abbreviations: GNS, GlcNAcstatin C; *O*-GlcNAc, *O*-linked *N*-acetylglucosamine.

cytokine LIF. However, we could not detect any change in Stat3 phosphorylation in GNS-treated cells compared to controls (Fig. 3A), which suggests that the effect of raised *O*-GlcNAc levels on ESC differentiation is independent of Stat3 activity. Furthermore, GNS is not able to substitute for LIF and promotes the clonal expansion of ESCs in serum (Supporting Information Fig. S3A), nor does it affect cloning efficiency in the presence of LIF (Fig. 3B), consistent with a delay rather than a complete block in differentiation.

O-GlcNAc Levels Do Not Affect Transition of Naïve Cells to a Primed State

Recent work has suggested that *O*-GlcNAcylation of Oct4 can maintain ESCs by regulation of downstream naïve state markers like *Nanog*, *Rex1*, *Klf2*, *Klf5* and others [11]. When cultured in the presence of serum and LIF, ESC populations consist of mixed naïve and primed cells, but in serum-free N2B27 monolayer differentiation conditions the transcripts for the naïve markers decline sharply after ~24 hours and the cells proceed to differentiate. To test whether this early transition is affected by *O*-GlcNAc levels we used the OCRG9 line, expressing GFP under the control of the endogenous *Rex1* locus. Under control conditions, OCRG9 cells become GFP negative at day 2 of differentiation (the delay between loss of *Rex1* mRNA and loss of GFP is due to the stability of the latter). GNS-treated OCRG9 cells lost expression of *Rex1*-GFP at the same rate as the control (DMSO-treated) cells (Fig. 3C). The transition from naïve to primed state is mediated by the actions of the Erk1/2 kinases downstream of fibroblast growth factor (Fgf) signaling. Recent work in *Drosophila* has identified

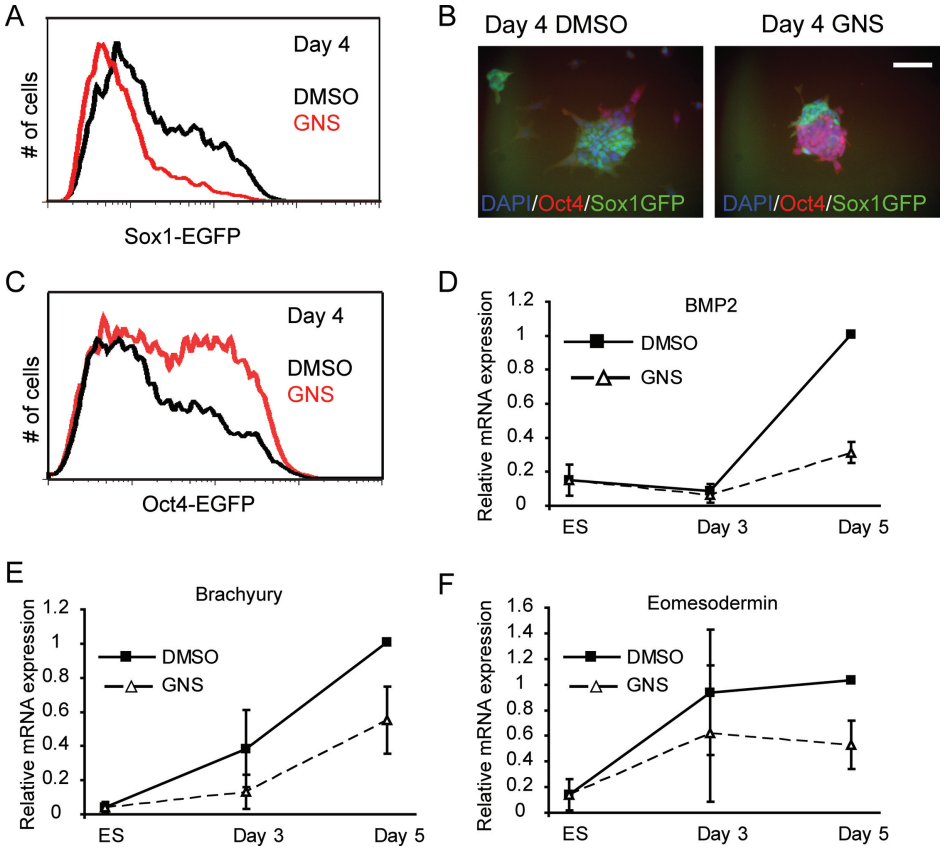

Figure 2. GlcNAcstatin treatment impairs embryonic stem cell (ESC) differentiation. **(A):** The proportion of Sox1GFP expressing neural progenitors is decreased following GNS treatment during differentiation of 46C cells. **(B):** 46C cells retain expression of Oct4 when differentiated in the presence of GNS while control cells differentiate into neural progenitors. Scale bar = 50 μm. **(C):** Oct4GiP cells retain Oct4-GFP expression at higher levels in differentiation conditions compared to vehicle only controls. **(D–F):** ESCs were differentiated as embryoid bodies in the presence of DMSO or GNS for 3 or 5 days then assayed for expression of endodermal marker BMP2 (D) or mesodermal markers Brachyury (E) or Eomesodermin (F) by RT-qPCR. Mean ± SEM from $n = 3$ experiments. Abbreviations: DAPI, 4′,6-diamidino-2-phenylindole; GFP, green fluorescent protein; GNS, GlcNAcstatin C.

a requirement for *O*-GlcNAcylation for Erk1/2 signal transduction downstream of Fgf receptor [30]. However, consistent with a normal transition to the primed state, we did not detect any changes in the profile of Erk1/2 phosphorylation in GNS-treated cells (Supporting Information Fig. S3B), suggesting that the signal regulating transition of naïve cells to primed pluripotency is unaltered by increased *O*-GlcNAc levels.

This is further confirmed by quantitative RT-PCR for other naïve cell markers *Nanog*, *Rex1/Zfp42*, *Esrrb*, and *Dppa3* (Fig. 3D, Supporting Information Fig. S3C) as well as the transient upregulation of the epiblast marker Fgf5 in EB differentiation (Fig. 3E). Taken together, these results indicate that *O*-GlcNAc levels do not affect the initial events of ESC differentiation.

In order to test this hypothesis further we performed a microarray experiment to measure GNS's effect on global transcriptional changes during the transition of naïve pluripotent cells to a primed state. We treated ESCs with DMSO or GNS for 24 hours either in ESC media or in neural differentiation conditions and compared gene expression between treatments. Principal component analysis showed that the change from self-renewal media to monolayer differentiation media (including serum withdrawal) causes a major effect on global gene expression (37%). This is reflected in the clustering of the most significantly regulated genes that change their expression from ESCs to day 1 of differentiation irrespective of GNS (Fig. 3F). We then compared the changes in gene expression during the first 24 hours of differentiation in the presence of GNS or DMSO compared to undifferentiated ESCs. This revealed that a great majority of the genes are common to the two conditions (Fig. 3G), consistent with the hypothesis that naïve-to-primed transition is unaffected by *O*-GlcNAc levels. Many of the genes downregulated (fold change >1.2, $p < .05$) in both vehicle and GNS samples belong to pathways regulating ESC self-renewal (i.e., MAPK, Jak/Stat and Tgfβ superfamily; Table 1) and include *Bmp4*, *Tgfβ1* and downstream targets *Id1*, *Id2*, *Id3*; *Stat3* and downstream target *Socs3* as well as transcription factors associated with naïve pluripotency (*Klf2*, *3*, *4*, and *5*, *Nanog*, *Rex1/Zfp42*). Common upregulated genes include neural and mesodermal early differentiation regulators (*Otx2*, *Tbx4*, *Neurogenin3*, *NeuroD1*, and *NeuroD4*) as well as primed pluripotency epiblast marker *Fgf5*, consistent with our RT-qPCR results. In total, 45 genes from our set also belong to the Plurinet network of pluripotency-related markers [31], 20 of which are

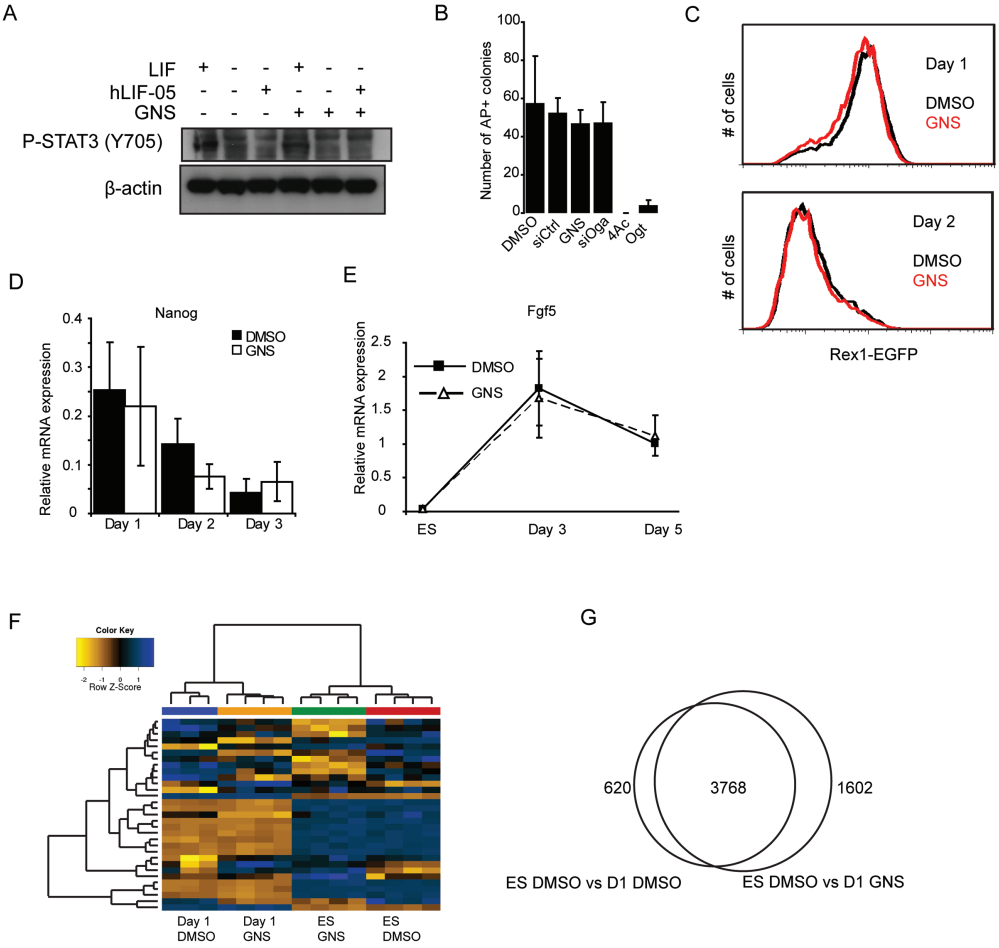

Figure 3. Increased *O*-GlcNAcylation does not affect naïve-primed embryonic stem cell (ESC) transition. **(A):** 46C ESCs were cultured in the absence of LIF overnight and then stimulated with either LIF or the LIF receptor antagonist hLIF-05 in the presence of either DMSO or GNS, then blotted for phospho-Stat3 (Y705). GNS treatment does not interfere with the ability of LIF to stimulate Stat3 phosphorylation and does not cause Stat3 phosphorylation in the absence of LIF. **(B):** GNS or Oga knockdown does not affect the ability of ESCs to form undifferentiated colonies, while inhibition or knockdown of Ogt results in cell death. 4Ac: 4-Acetyl-5S-GlcNAc (Ogt inhibitor). **(C):** OCRG9 ESCs (Rex1EGFP reporter line) were differentiated in N2B27 for 2 days in vehicle or GNS and analyzed by flow cytometry for EGFP fluorescence. GlcNAcstatin treatment does not affect the loss of naïve marker Rex1. **(D):** Quantitative RT-PCR analysis during monolayer differentiation shows that expression of the naïve marker Nanog is lost at equivalent rates in the presence of vehicle (DMSO; black bars) or GNS (white bars). **(E):** Upregulation of the primed pluripotency marker Fgf5 is unaffected by GNS treatment (D and E, $n = 3$, mean $\pm$ SEM). **(F):** Hierarchical clustering (using Euclidean distance with complete linkage agglomeration method) of samples based on the top 31 differentially expressed genes (ranked by *p*-value) separates the samples according to the stage of differentiation and shows clustering of genes that contribute to differentiation. **(G):** Pairwise comparison of genes regulated during the first 24 hours of differentiation in vehicle (DMSO) or GNS (fold change >1.2, $p < .05$). Abbreviation: GNS, GlcNAcstatin C.

upregulated in ESCs, and 25 are upregulated in the day 1 samples (Supporting Information Tables S3 and S4, respectively). This result confirms that raised *O*-GlcNAc levels do not affect the expression of genes controlling the onset of differentiation in serum-free monolayer.

Increased Protein *O*-GlcNAcylation Does Not Interfere With Phosphorylation of Major Signaling Pathways in ESCs

O-GlcNAc is frequently attached to serine and threonine residues that can also be phosphorylated and for many proteins this reciprocal relationship (often described as a "Yin-Yang") regulates their activity. To investigate whether increased *O*-GlcNAc levels affect the global phosphorylation of proteins in ESCs we compared the global phospho-serine and phospho-threonine levels of ESC protein extracts treated with various concentrations of GNS. Unlike global *O*-GlcNAc levels that showed a dose-dependent increase, protein phosphorylation levels remained relatively constant, suggesting that the majority of ESC phosphosites are not occupied by *O*-GlcNAc (Supporting Information Fig. S4A). This result, however, could reflect a high abundance of constitutively phosphorylated proteins in these cells that conceal significant changes in less abundant and more dynamically regulated phosphoproteins. We therefore focused our attention on protein kinases as these proteins are often dynamically regulated by phosphorylation and mediate numerous important cellular responses. We analyzed kinase phosphorylation in whole-cell lysates from ESCs treated for 24 hours with either DMSO or GNS using a protein kinase array. However, none of the 46 kinase sites

Table 1. KEGG pathway analysis of the genes significantly regulated during the first 24 hours of embryonic stem cell differentiation irrespective of GlcNAcstatin C treatment

Regulation	Term	Count	% of genes in group	*p*-Value	Benjamini
Up in ES	Focal adhesion	57	0.3	2.30E-12	3.90E-10
Up in ES	Lysosome	39	0.2	1.90E-10	1.60E-08
Up in ES	ECM-receptor interaction	28	0.2	5.70E-08	3.10E-06
Up in ES	Regulation of actin cytoskeleton	45	0.2	1.70E-05	7.20E-04
Up in ES	TGF-beta signaling pathway	24	0.1	2.80E-05	9.20E-04
Up in ES	Pathways in cancer	59	0.3	3.90E-05	1.10E-03
Up in ES	Tight junction	31	0.2	6.70E-05	1.60E-03
Up in ES	Leukocyte transendothelial migration	28	0.2	1.10E-04	2.20E-03
Up in ES	MAPK signaling pathway	49	0.3	1.50E-04	2.70E-03
Up in ES	ARVC	20	0.1	2.50E-04	4.20E-03
Up in ES	Small cell lung cancer	21	0.1	5.00E-04	7.50E-03
Up in ES	Other glycan degradation	8	<0.1	7.50E-04	1.00E-02
Up in ES	Glycosphingolipid biosynthesis	7	<0.1	2.10E-03	2.60E-02
Up in ES	Endocytosis	36	0.2	2.60E-03	3.00E-02
Up in ES	Jak-STAT signaling pathway	29	0.2	2.70E-03	3.00E-02
Up in ES	HCM	19	0.1	2.90E-03	3.00E-02
Up in ES	Glycosphingolipid biosynthesis	7	<0.1	3.20E-03	3.10E-02
Up in ES	ErbB signaling pathway	19	0.1	4.30E-03	3.90E-02
Up in ES	Amino sugar and nucleotide sugar metabolism	12	0.1	5.50E-03	4.70E-02
Up in ES	Galactose metabolism	9	<0.1	5.80E-03	4.70E-02
Up in ES	Dorso-ventral axis formation	8	<0.1	6.30E-03	4.90E-02
Up in Day 1	Steroid biosynthesis	11	0.1	1.90E-08	3.40E-06
Up in Day 1	Terpenoid backbone biosynthesis	8	0.1	1.30E-05	1.10E-03
Up in Day 1	Biosynthesis of unsaturated fatty acids	9	0.1	2.60E-04	1.50E-02
Up in Day 1	Pathways in cancer	38	0.3	6.90E-04	3.00E-02

Abbreviations: ARVC, arrhythmogenic right ventricular cardiomyopathy; HCM, hypertrophic cardiomyopathy.

profiled showed a significant change in their basal level of phosphorylation by GNS (Supporting Information Fig. S4B) indicating that the ESC kinome is not significantly affected by the Yin-Yang interplay between *O*-GlcNAc and phosphate.

Raised *O*-GlcNAc Levels Affect Transcription of Repressed Genes in ESCs

We then analyzed ESCs treated for 24 hours with DMSO or GNS by microarrays, using Ranked Product analysis—a method better at detecting changes in gene expression from small number of replicates than the more commonly used Significance Analysis of Microarrays or Analysis of Variance methods [32]. Using a Ranked Product p-value $< .05$ we identified 971 differentially regulated genes. Of these, 516 were upregulated following GNS treatment and 455 were downregulated. Two of the most significantly regulated genes are *Ogt* and *Oga* (Fig. 4A). Using RT-qPCR we found that *Ogt* expression is significantly reduced within 1 hour of treatment with GNS, whereas *Oga* upregulation appears slower (Fig. 4B, 4C). The regulation of Oga and Ogt by *O*-GlcNAc levels is also reflected at the protein level within 24 hours of treatment with GNS (Supporting Information Fig. S4C) or by knockdown of Ogt or Oga using siRNA (Fig. 4D).

One other gene that stood out from this analysis is Zscan4 (Fig. 4A). This gene has previously been associated with telomere maintenance in ESCs [33] and is associated with a subpopulation of cells similar to that of the recently described "2C" state of privileged developmental plasticity, existing within ESC cultures [34–36]. 2C cells differ in gene expression from ESCs in that they express genes associated with zygotic genome activation and have been demonstrated to be totipotent (giving rise to extraembryonic as well as embryonic tissues in chimeras) [36]. Transcripts marking this subpopulation include retrotransposons normally repressed by epigenetic mechanisms as well as chimeric transcripts of genes with junctions to murine endogenous retrovirus with leucine tRNA primer (MERVL) elements [36]. Interestingly, the number of genes upregulated in the GNS-treated samples was much larger than the DMSO samples, both for the genes enriched in ESCs and for genes enriched in Day 1 differentiating cells. We therefore performed pairwise comparisons between the ESC and Day 1 samples in DMSO or GNS treatment for the most regulated genes ($p < .05$, fold change >2). The majority of genes with expression higher in ESCs than Day 1 ($\sim$80%; 348/437) were similarly regulated both in DMSO and GNS. However, of the remaining 20% (those that were not common to GNS and DMSO), nearly three times more genes were higher in the GNS sample than in DMSO (65 vs. 24; Fig. 5A). Similarly, the majority of the genes expressed at higher level in Day 1 samples compared to ESCs are common to GNS and DMSO ($\sim$75%; 160/214), but of those differentially regulated between the treatments, those regulated by GNS outnumbered those regulated by DMSO by a factor of 3.5 (42 vs. 12; Fig. 5A). This result suggests that GNS treatment causes a general increase in gene expression both in ESCs and early differentiating cells (Fig. 3G). We then turned to GSEA for further mining of our expression data. Search of the whole MSigDB and GeneSigDB databases using GSEA did not reveal a significant enrichment for any gene set so we created our own gene sets from publicly available microarray and ChIP-seq data as published in relevant papers. We focused our attention to targets of the ESC core pluripotency network [37, 38] and genes regulated by the Polycomb group [39–41]. Although there was no significant enrichment for polycomb complex 1 or 2 targets in our dataset (Supporting Information Table S6), a specific subset of polycomb target genes, designated as Polycomb-repressed (PRCR) [39], is highly

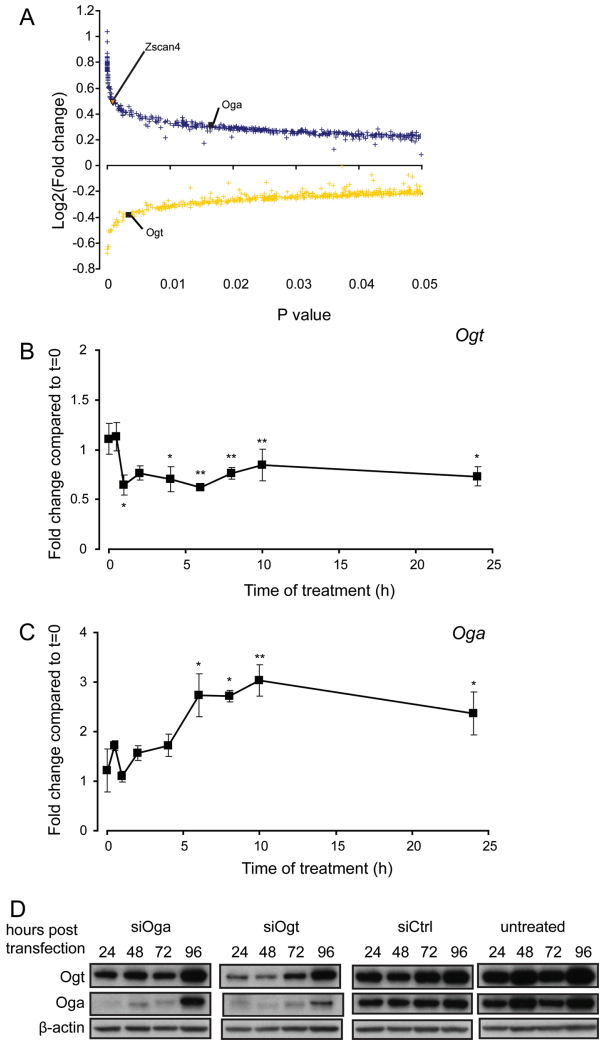

Figure 4. *O*-linked *N*-acetylglucosamine levels control Ogt and Oga levels in embryonic stem cells (ESCs). **(A):** Ranked product analysis of genes regulated in ESCs by GlcNAcstatin C treatment. Blue: significantly upregulated genes; yellow: significantly downregulated genes. **(B):** Validation of *Ogt* regulation by RT-qPCR ($n = 4$, mean ± SEM). **(C):** Validation of *Oga* expression by RT-qPCR ($n = 4$, mean ± SEM). **(D):** RNAi knockdown of Ogt and Oga results in a negative feedback regulation on the expression of Oga and Ogt, respectively. *, $p < .05$; **, $p < .01$ compared to $t = 0$ (t test).

significantly enriched in ESCs treated with GNS compared to the vehicle control (Fig. 5B, Table 2). Strikingly, we also detected highly significant enrichment in the GNS-treated samples for the gene sets corresponding to the 2C population (the 2C sets as described previously [36]) as well as the gene set "2C-MERVL+" which comprises the genes from the 2C set that contain MERVL chimeric transcripts. The enrichment of 2C transcripts is highly significant both in undifferentiated ESCs and in Day 1 differentiation samples treated with GNS compared to their respective vehicle controls (Fig. 5C, 5D, Table 2, Supporting Information Tables S6 and S7). We validated the upregulation of three 2C associated genes (Zfp352, Tdpoz3, and Zscan4) by Oga knockdown and GNS in ESCs using quantitative RT-PCR and show a consistent upregulation for all three (Fig. 5C, Supporting Information Fig. S5A–S5C). These results therefore indicate that increased *O*-GlcNAc levels result in increased expression of epigenetically repressed genes, including those characteristic of a totipotent 2C state.

Discussion

Our results demonstrate that ESC differentiation is reduced under conditions of increased *O*-GlcNAc signaling and pinpoint the deficit to a stage in differentiation after the loss of naive pluripotency but before the definitive differentiation as marked by the loss of Oct4 expression. Using a range of differentiation protocols we found a delay in the onset of differentiation toward embryonic lineages. We report that Oga inhibition does not affect the upregulation or downregulation of the genes known to be regulated early during differentiation, but results in prolonged expression of the pluripotency master regulator Oct4 and a delay in the acquisition of stable neural differentiation marker Sox1. Therefore, our results clearly demonstrate that Oga inhibition does not affect the transition from the naïve to the primed state, indicating instead that elevated *O*-GlcNAc levels affect differentiation progression at a later stage. Taken together, our data show a clear effect of Oga inhibition on the initiation of ESC differentiation that affects multiple somatic lineages.

We also report that Oga inhibition does not interfere with the steady-state phosphorylation levels of several key kinases, suggesting that the main mechanism of action operates downstream of the signal transduction machinery. The fact that the phosphorylation status of kinases is unaffected by GNS in this context does not, however, preclude changes in kinase activity, localization, or interaction with substrates. It also remains possible that the stimulated, maximal phosphorylation levels of some of these kinases will be affected by elevated *O*-GlcNAc despite no obvious effects at the basal phosphorylation levels.

An early study of Oga inhibition during ESC differentiation identified a deficit in spontaneous cardiomyocyte generation in EBs [12]. In that study a different, less specific Oga inhibitor was added to 5-day-old differentiating EBs, when ESC markers are already lost and the cells have already committed to differentiating, highlighting a role for Oga in the commitment of mesodermal cells to the cardiomyocyte lineage. Our data extend and refine this previous observation by determining an earlier role in the exit from the pluripotent state. More recently, a report demonstrated a role for the *O*-GlcNAc modification of Oct4 in the onset of ESC differentiation [11]. Our results are consistent with those findings, although we find that naïve ESC markers decline normally under elevated *O*-GlcNAc conditions, demonstrating that the earliest events in ESC differentiation are not *O*-GlcNAc dependent.

A number of recent papers have linked the *O*-GlcNAc transferase to the regulation of epigenetic regulators [16–18, 42]. Our gene expression analysis has revealed that following Oga inhibition a number of repressed genes become activated. Gene set enrichment analysis showed a significant enrichment for a specific subset of polycomb repressor complex target genes, as well as genes associated with the early zygotic genome activation and retrovirus-like elements. This latter gene expression signature is thought to mark a developmentally privileged ESC subpopulation capable of differentiation into extraembryonic lineages [35, 36]. These genes are

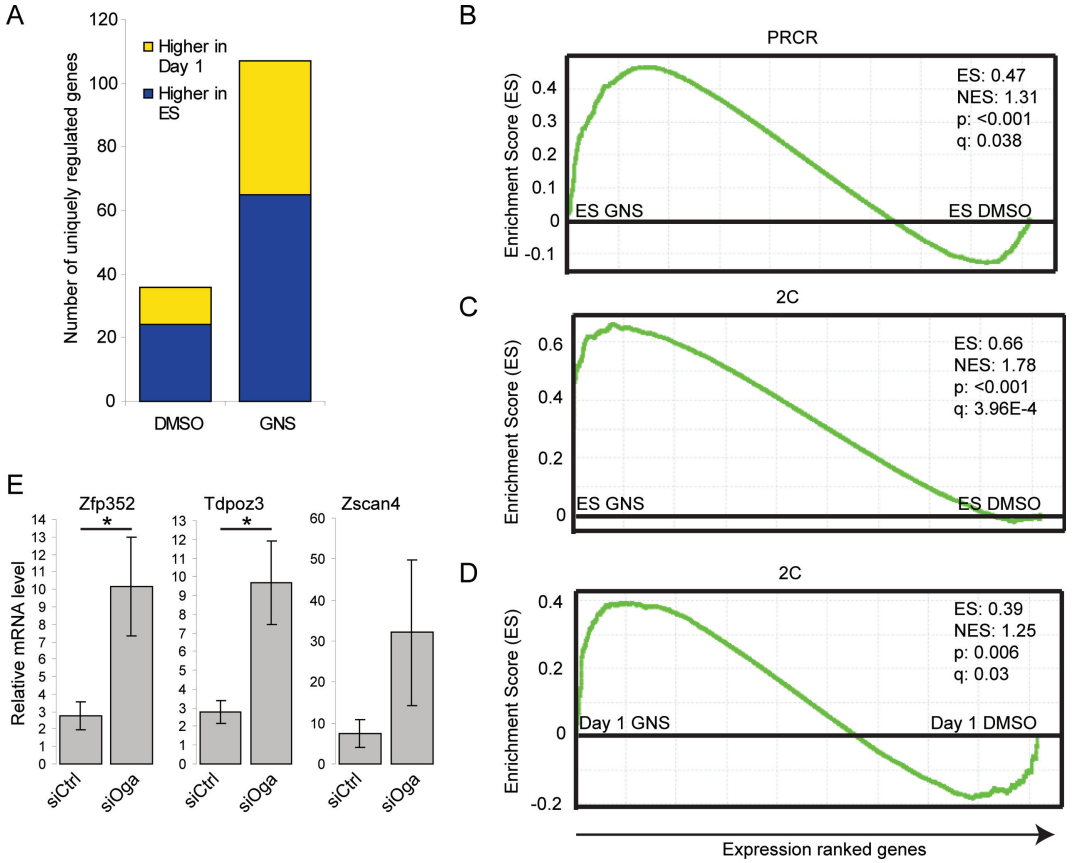

Figure 5. Oga inhibition results in a broad genomic transcriptional derepression. **(A):** Bar charts showing the number of unique genes upregulated in ESCs or in Day 1 differentiation in the DMSO or GNS samples ($p < .05$, fold change >2). **(B):** Gene set enrichment analysis (GSEA) shows that GNS treatment results in an upregulation of the "Polycomb repressed" gene set. **(C):** GSEA showing enrichment of the 2C gene set in the GNS-treated ESCs. **(D):** The 2C gene set is also enriched in the GNS-treated Day 1 differentiating cells. **(E):** RT-qPCR validation of the upregulation of three genes from the 2C cohort by knockdown of Oga ($n = 3$, mean ± SEM) *, $p < .05$. Abbreviation: GNS, GlcNAcstatin C.

Table 2. Gene sets significantly enriched in GlcNAcstatin C or DMSO-treated embryonic stem cell or Day 1 differentiation samples

Name	Size	Enrichment score	Normalized enrichment score	Nominal *p*-value	False discovery rate *q*-value	Enrichment in
2C-MERVL+	45	0.85	1.85	<.001	<0.001	ES GNS versus ES DMSO
2C	466	0.66	1.78	<.001	0.001	ES GNS versus ES DMSO
PRCR	1,331	0.47	1.31	<.001	0.038	ES GNS versus ES DMSO
2C	466	0.39	1.25	.006	0.033	Day 1 GNS versus Day 1 DMSO

under the control of epigenetic regulators of gene expression and their expression can be induced by interfering with histone acetylation and methylation, either pharmacologically or by genetic deletion of key enzymes [36]. Our findings therefore suggest that the levels of *O*-GlcNAc in cells influence global gene expression regulated by major epigenetic mechanisms such as histone modifications and DNA methylation, and are consistent with a differentiation delay observed in ESCs treated with an inhibitor of histone deacetylases [43]. Intriguingly, Oga has been shown to possess histone acetyltransferase (HAT) activity in vitro [44] and this activity is distinct from its GlcNAcase function. It is therefore likely that although inhibition of Oga's *O*-GlcNAcase activity by GNS raises *O*-GlcNAc levels and leads to feedback Oga upregulation, it does not affect the protein's HAT activity (as has been shown for the less specific inhibitor Streptozotocin [44]). This could possibly lead to increased histone acetylation and the derepression of transcription we observe, although the observation that 2C genes were upregulated even when Oga protein levels are reduced by siRNA (Fig. 5E) indicates that this is not the main mechanism in this case. The negative feedback regulation from increased *O*-GlcNAc levels to a reduction in Ogt protein levels may also contribute to the phenotypes observed, by disrupting the formation of some of its protein complexes. *O*-GlcNAc can itself modify histones [45] and promote transcriptional activation [46], so the precise mechanism by which gene expression is regulated by *O*-GlcNAc levels is likely to be complex and involve many, possibly redundant molecular players.

Conclusions

In conclusion, our results demonstrate that ESC differentiation is delayed under conditions of increased *O*-GlcNAc signaling and pinpoint the delay to a stage in differentiation after the loss of naive pluripotency but before commitment to differentiation as marked by the loss of Oct4 expression. We used defined media and short timescales to minimize the possibility of multiple effects on several interacting cell types, and found a measurable, significant defect in the progression toward differentiation. We attribute this differentiation defect to a disruption of the normal function of global activation and repression complexes, such as the histone deacetylases, polycomb group, and Ten-eleven translocation proteins as well as direct effects of Oga and Ogt on chromatin.

Acknowledgments

This work was funded by Tenovus Tayside small project grants (M.P.S.), a Ninewells Cancer Campaign studentship (T.C.E.D.), a Thai Government scholarship (W.W.) and by a Wellcome Trust Senior Research Fellowship (Grant Number WT087590MA) to D.M.F.v.A. We would like to thank Ian Chambers for hLIF-05, Osama Albarbarawi for mass spectrometry, Pauline Gammack, Claudio Guerri, Claire Higgins, and Panagiotis Stamatiadis for technical assistance and Arno Müller, Steve Keyse, and Kate Storey for helpful comments on the manuscript.

Author Contributions

C.M.S., T.C.E.D., and W.W.: collection and assembly of data and data analysis; K.S.: experimental design, collection and assembly of data, and data analysis and interpretation; M.M. and G.J.B.: data analysis and interpretation; D.M.F.v.A.: provision of reagents, experimental design, and data analysis and interpretation; M.P.S.: conception and design, data collection and assembly of data, data analysis and interpretation, manuscript writing, and final approval of manuscript.

Disclosure of Potential Conflicts of Interest

The authors indicate no potential conflicts of interest.

References

1 Stavridis MP. Embryonic stem cells: A signalling perspective. In: St John JC, ed. Mitochondrial DNA, Disease and Stem Cells. 1st ed. New York: Humana Press, 2012.
2 Nichols J, Smith A. Naive and primed pluripotent states. Cell Stem Cell 2009;4:487–492.
3 Torres CR, Hart GW. Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc. J Biol Chem 1984;259:3308–3317.
4 Dias WB, Hart GW. O-GlcNAc modification in diabetes and Alzheimer's disease. Mol Biosyst 2007;3:766–772.
5 Shafi R, Iyer SP, Ellies LG, et al. The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny. Proc Natl Acad Sci USA 2000;97:5735–5739.
6 O'Donnell N, Zachara NE, Hart GW, et al. Ogt-dependent X-chromosome-linked protein glycosylation is a requisite modification in somatic cell function and embryo viability. Mol Cell Biol 2004;24:1680–1690.
7 van den Berg DL, Snoek T, Mullin NP, et al. An Oct4-centered protein interaction network in embryonic stem cells. Cell Stem Cell 2010;6:369–381.
8 Pardo M, Lang B, Yu L, et al. An expanded Oct4 interaction network: Implications for stem cell biology, development, and disease. Cell Stem Cell 2010;6:382–395.
9 Ding J, Xu H, Faiola F, et al. Oct4 links multiple epigenetic pathways to the pluripotency network. Cell Res 2012;22:155–167.
10 Webster DM, Teo CF, Sun Y, et al. O-GlcNAc modifications regulate cell survival and epiboly during zebrafish development. BMC Dev Biol 2009;9:28.
11 Jang H, Kim TW, Yoon S, et al. O-GlcNAc regulates pluripotency and reprogramming by directly acting on core components of the pluripotency network. Cell Stem Cell 2012;11:62–74.
12 Kim HS, Park SY, Choi YR, et al. Excessive O-GlcNAcylation of proteins suppresses spontaneous cardiogenesis in ES cells. FEBS Lett 2009;583:2474–2478.
13 Sinclair DA, Syrzycka M, Macauley MS, et al. Drosophila O-GlcNAc transferase (OGT) is encoded by the Polycomb group (PcG) gene, super sex combs (sxc). Proc Natl Acad Sci USA 2009;106:13427–13432.
14 Gambetta MC, Oktaba K, Muller J. Essential role of the glycosyltransferase sxc/Ogt in polycomb repression. Science 2009;325:93–96.
15 Yang X, Zhang F, Kudlow JE. Recruitment of O-GlcNAc transferase to promoters by corepressor mSin3A: Coupling protein O-GlcNAcylation to transcriptional repression. Cell 2002;110:69–80.
16 Vella P, Scelfo A, Jammula S, et al. Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. Mol Cell 2013;49:645–656.
17 Shi F, Kim H, Lu W, et al. Ten-eleven translocation 1 (Tet1) is regulated by O-GlcNAc transferase (Ogt) for target gene repression in mouse embryonic stem cells. J Biol Chem 2013;288:20776–20784.
18 Chen Q, Chen Y, Bian C, et al. TET2 promotes histone O-GlcNAcylation during gene transcription. Nature 2013;493:561–564.
19 Stavridis MP, Lunn JS, Collins BJ, et al. A discrete period of FGF-induced Erk1/2 signalling is required for vertebrate neural specification. Development 2007;134:2889–2894.
20 Ying QL, Stavridis M, Griffiths D, et al. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. Nat Biotechnol 2003;21:183–186.
21 Aubert J, Stavridis MP, Tweedie S, et al. Screening for mammalian neural genes via fluorescence-activated cell sorter purification of neural precursors from Sox1-gfp knock-in mice. Proc Natl Acad Sci USA 2003;100:11836–11841.
22 Toyooka Y, Shimosato D, Murakami K, et al. Identification and characterization of subpopulations in undifferentiated ES cell culture. Development 2008;135:909–918.
23 Ying QL, Nichols J, Evans EP, et al. Changing potency by spontaneous fusion. Nature 2002;416:545–548.
24 Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 2001;29:e45.
25 Smyth GK. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 2004;3:Article3.
26 Hong F, Breitling R, McEntee CW, et al. RankProd: A bioconductor package for detecting differentially expressed genes in meta-analysis. Bioinformatics 2006;22:2825–2827.
27 MacKenzie KF, Clark K, Naqvi S, et al. PGE(2) induces macrophage IL-10 production and a regulatory-like phenotype via a protein kinase A-SIK-CRTC3 pathway. J Immunol 2013;190:565–577.
28 Dorfmueller HC, Borodkin VS, Schimpl M, et al. GlcNAcstatins are nanomolar inhibitors of human O-GlcNAcase inducing cellular hyper-O-GlcNAcylation. Biochem J 2009;420:221–227.
29 Gloster TM, Zandberg WF, Heinonen JE, et al. Hijacking a biosynthetic pathway yields a glycosyltransferase inhibitor within cells. Nat Chem Biol 2011;7:174–181.
30 Mariappa D, Sauert K, Marino K, et al. Protein O-GlcNAcylation is required for fibroblast growth factor signaling in Drosophila. Sci Signal 2011;4:ra89.
31 Muller FJ, Laurent LC, Kostka D, et al. Regulatory networks define phenotypic classes of human stem cell lines. Nature 2008;455:401–405.

32 Jeffery IB, Higgins DG, Culhane AC. Comparison and evaluation of methods for generating differentially expressed gene lists from microarray data. BMC Bioinformatics 2006;7:359.
33 Zalzman M, Falco G, Sharova LV, et al. Zscan4 regulates telomere elongation and genomic stability in ES cells. Nature 2010;464:858–863.
34 Amano T, Hirata T, Falco G, et al. Zscan4 restores the developmental potency of embryonic stem cells. Nat Commun 2013;4:1966.
35 Morgani SM, Canham MA, Nichols J, et al. Totipotent embryonic stem cells arise in ground-state culture conditions. Cell Rep 2013;3:1945–1957.
36 Macfarlan TS, Gifford WD, Driscoll S, et al. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature 2012;487:57–63.
37 Kim J, Chu J, Shen X, et al. An extended transcriptional network for pluripotency of embryonic stem cells. Cell 2008;132:1049–1061.
38 Chen X, Xu H, Yuan P, et al. Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell 2008;133:1106–1117.
39 Brookes E, de Santiago I, Hebenstreit D, et al. Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs. Cell Stem Cell 2012;10:157–170.
40 Mikkelsen TS, Ku M, Jaffe DB, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 2007;448:553–560.
41 Ku M, Koche RP, Rheinbay E, et al. Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet 2008;4:e1000242.
42 Deplus R, Delatte B, Schwinn MK, et al. TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. EMBO J 2013;32:645–655.
43 Ware CB, Wang L, Mecham BH, et al. Histone deacetylase inhibition elicits an evolutionarily conserved self-renewal program in embryonic stem cells. Cell Stem Cell 2009;4:359–369.
44 Toleman C, Paterson AJ, Whisenhunt TR, et al. Characterization of the histone acetyltransferase (HAT) domain of a bifunctional protein with activable O-GlcNAcase and HAT activities. J Biol Chem 2004;279:53665–53673.
45 Sakabe K, Wang Z, Hart GW. {beta}-N-acetylglucosamine (O-GlcNAc) is part of the histone code. Proc Natl Acad Sci USA 2010;107:19915–19920.
46 Fujiki R, Hashiba W, Sekine H, et al. GlcNAcylation of histone H2B facilitates its monoubiquitination. Nature 2011;480:557–560.

See www.StemCells.com for supporting information available online.

Contributions:

- Established siOga/siOgt knockdown protocol
- Experiments and data analysis of the following results:
  - Figure 2: D, E and F
  - Figure 3: E
  - Figure 4: D
  - Figure 5: E
- Comments and suggestions before submission

B. References

Aiken CEM, Swoboda PPL, Skepper JN, Johnson MH (2004) The direct measurement of embryogenic volume and nucleo-cytoplasmic ratio during mouse pre-implantation development. *Reproduction (Cambridge, England)*, **128**(5): 527–535

Aloia L, Di Stefano B, Di Croce L (2013) Polycomb complexes in stem cells and embryonic development. *Development (Cambridge, England)*, **140**(12): 2525–2534

Andrews S (2010) FastQC A Quality Control tool for High Throughput Sequence Data. *http://www.bioinformatics.babraham.ac.uk/projects/fastqc*

Aubert J, Dunstan H, Chambers I, Smith A (2002) Functional gene screening in embryonic stem cells implicates Wnt antagonism in neural differentiation. *Nature biotechnology*, **20**(12): 1240–1245

Bauer C, Gobel K, Nagaraj N, Colantuoni C, Wang M, Muller U, Kremmer E, Rottach A, Leonhardt H (2015) Phosphorylation of TET proteins is regulated via O-GlcNAcylation by the O-linked N-acetylglucosamine transferase (OGT). *The Journal of biological chemistry*, **290**(8): 4801–4812

Beausoleil SA, Jedrychowski M, Schwartz D, Elias JE, Villen J, Li J, Cohn MA, Cantley LC, Gygi SP (2004) Large-scale characterization of HeLa cell nuclear phosphoproteins. *Proceedings of the National Academy of Sciences of the United States of America*, **101**(33): 12130–12135

Bensaude O, Babinet C, Morange M, Jacob F (1983) Heat shock proteins, first major products of zygotic gene activity in mouse embryo. *Nature*, **305**(5932): 331–333

Bock C, Reither S, Mikeska T, Paulsen M, Walter J, Lengauer T (2005) BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing. *Bioinformatics (Oxford, England)*, **21**(21): 4067–4068

Boeuf H, Hauss C, Graeve FD, Baran N, Kedinger C (1997) Leukemia inhibitory factor-dependent transcriptional activation in embryonic stem cells. *The Journal of cell biology*, **138**(6): 1207–1217

Boland MJ, Hazen JL, Nazor KL, Rodriguez AR, Gifford W, Martin G, Kupriyanov S, Baldwin KK (2009) Adult mice generated from induced pluripotent stem cells. *Nature*, **461**(7260): 91–94

Boroviak T, Nichols J (2014) The birth of embryonic pluripotency. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **369**(1657)

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry*, **72**: 248–254

Brambrink T, Foreman R, Welstead GG, Lengner CJ, Wernig M, Suh H, Jaenisch R (2008) Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. *Cell stem cell*, **2**(2): 151–159

Briscoe J, Small S (2015) Morphogen rules: design principles of gradient-mediated embryo patterning. *Development (Cambridge, England)*, **142**(23): 3996–4009

Brons IGM, Smithers LE, Trotter MWB, Rugg-Gunn P, Sun B, Chuva de Sousa Lopes, Susana M, Howlett SK, Clarkson A, Ahrlund-Richter L, Pedersen RA, Vallier L (2007) Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature*, **448**(7150): 191–195

Brookes E, de Santiago I, Hebenstreit D, Morris KJ, Carroll T, Xie SQ, Stock JK, Heidemann M, Eick D, Nozaki N, Kimura H, Ragoussis J, Teichmann SA, Pombo A (2012) Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs. *Cell stem cell*, **10**(2): 157–170

Buganim Y, Faddah DA, Cheng AW, Itskovich E, Markoulaki S, Ganz K, Klemm SL, van Oudenaarden A, Jaenisch R (2012) Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell*, **150**(6): 1209–1222

Burridge PW, Matsa E, Shukla P, Lin ZC, Churko JM, Ebert AD, Lan F, Diecke S, Huber B, Mordwinkin NM, Plews JR, Abilez OJ, Cui B, Gold JD, Wu JC (2014) Chemically defined generation of human cardiomyocytes. *Nature methods*, **11**(8): 855–860

Butkinaree C, Cheung WD, Park S, Park K, Barber M, Hart GW (2008) Characterization of beta-N-acetylglucosaminidase cleavage by caspase-3 during apoptosis. *The Journal of biological chemistry*, **283**(35): 23557–23566

Chambers SM, Fasano CA, Papapetrou EP, Tomishima M, Sadelain M, Studer L (2009) Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nature biotechnology*, **27**(3): 275–280

Chan YS, Goke J, Ng JH, Lu X, Gonzales KAU, Tan CP, Tng WQ, Hong ZZ, Lim YS, Ng HH (2013) Induction of a human pluripotent state with distinct regulatory circuitry that resembles preimplantation epiblast. *Cell stem cell*, **13**(6): 663–675

Chari S, Mao S (2016) Timeline: iPSCs–The First Decade. *Cell stem cell*, **18**(2): 294

Chen Q, Chen Y, Bian C, Fujiki R, Yu X (2013) TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature*, **493**(7433): 561–564

Chin MH, Mason MJ, Xie W, Volinia S, Singer M, Peterson C, Ambartsumyan G, Aimiuwu O, Richter L, Zhang J, Khvorostov I, Ott V, Grunstein M, Lavon N, Benvenisty N, Croce CM, Clark AT, Baxter T, Pyle AD, Teitell MA, Pelegrini M, Plath K, Lowry WE (2009) Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell stem cell*, **5**(1): 111–123

Chopra VS, Hendrix DA, Core LJ, Tsui C, Lis JT, Levine M (2011) The polycomb group mutant esc leads to augmented levels of paused Pol II in the Drosophila embryo. *Molecular cell*, **42**(6): 837–844

Chou TY, Dang CV, Hart GW (1995) Glycosylation of the c-Myc transactivation domain. *Proceedings of the National Academy of Sciences of the United States of America*, **92**(10): 4417–4421

Comer FI, Hart GW (2001) Reciprocity between O-GlcNAc and O-phosphate on the carboxyl terminal domain of RNA polymerase II. *Biochemistry*, **40**(26): 7845–7852

Comer FI, Vosseller K, Wells L, Accavitti MA, Hart GW (2001) Characterization of a mouse monoclonal antibody specific for O-linked N-acetylglucosamine. *Analytical biochemistry*, **293**(2): 169–177

Comtesse N, Maldener E, Meese E (2001) Identification of a nuclear variant of MGEA5, a cytoplasmic hyaluronidase and a beta-N-acetylglucosaminidase. *Biochemical and biophysical research communications*, **283**(3): 634–640

Coucouvanis E, Martin GR (1995) Signals for death and survival: A two-step mechanism for cavitation in the vertebrate embryo. *Cell*, **83**(2): 279–287

Cuny GD, Yu PB, Laha JK, Xing X, Liu JF, Lai CS, Deng DY, Sachidanandan C, Bloch KD, Peterson RT (2008) Structure-activity relationship study of bone morphogenetic protein (BMP) signaling inhibitors. *Bioorganic & medicinal chemistry letters*, **18**(15): 4388–4392

D'Amour KA, Agulnick AD, Eliazer S, Kelly OG, Kroon E, Baetge EE (2005) Efficient differentiation of human embryonic stem cells to definitive endoderm. *Nature biotechnology*, **23**(12): 1534–1541

D'Amour KA, Bang AG, Eliazer S, Kelly OG, Agulnick AD, Smart NG, Moorman MA, Kroon E, Carpenter MK, Baetge EE (2006) Production of pancreatic hormone-expressing endocrine cells from human embryonic stem cells. *Nature biotechnology*, **24**(11): 1392–1401

Davidson KC, Mason EA, Pera MF (2015) The pluripotent state in mouse and human. *Development (Cambridge, England)*, **142**(18): 3090–3099

Deng J, Shoemaker R, Xie B, Gore A, LeProust EM, Antosiewicz-Bourget J, Egli D, Maherali N, Park IH, Yu J, Daley GQ, Eggan K, Hochedlinger K, Thomson J, Wang W, Gao Y, Zhang K (2009) Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nature biotechnology*, **27**(4): 353–360

Deplus R, Delatte B, Schwinn MK, Defrance M, Mendez J, Murphy N, Dawson MA, Volkmar M, Putmans P, Calonne E, Shih AH, Levine RL, Bernard O, Mercher T, Solary E, Urh M, Daniels DL, Fuks F (2013) TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *The EMBO journal*, **32**(5): 645–655

Derynck R (1994) TGF-beta-receptor-mediated signaling. *Trends in biochemical sciences*, **19**(12): 548–553

Dias WB, Hart GW (2007) O-GlcNAc modification in diabetes and Alzheimer's disease. *Molecular bioSystems*, **3**(11): 766–772

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, **29**(1): 15–21

Doi A, Park IH, Wen B, Murakami P, Aryee MJ, Irizarry R, Herb B, Ladd-Acosta C, Rho J, Loewer S, Miller J, Schlaeger T, Daley GQ, Feinberg AP (2009) Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nature genetics*, **41**(12): 1350–1353

Dorfmueller HC, Borodkin VS, Schimpl M, Shepherd SM, Shpiro NA, van Aalten DMF (2006) GlcNAcstatin: a picomolar, selective O-GlcNAcase inhibitor that modulates intracellular O-glcNAcylation levels. *Journal of the American Chemical Society*, **128**(51): 16484–16485

Dorfmueller HC, Borodkin VS, Schimpl M, van Aalten DMF (2009) GlcNAcstatins are nanomolar inhibitors of human O-GlcNAcase inducing cellular hyper-O-GlcNAcylation. *The Biochemical journal*, **420**(2): 221–227

Dorfmueller HC, Borodkin VS, Schimpl M, Zheng X, Kime R, Read KD, van Aalten DMF (2010) Cell-penetrant, nanomolar O-GlcNAcase inhibitors selective against lysosomal hexosaminidases. *Chemistry & biology*, **17**(11): 1250–1255

Dosch R, Gawantka V, Delius H, Blumenstock C, Niehrs C (1997) Bmp-4 acts as a morphogen in dorsoventral mesoderm patterning in Xenopus. *Development (Cambridge, England)*, **124**(12): 2325–2334

Durinck S, Spellman PT, Birney E, Huber W (2009) Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature protocols*, **4**(8): 1184–1191

Edwards RG, Gates AH (1959) Timing of the stages of the maturation divisions, ovulation, fertilization and the first cleavage of eggs of adult mice treated with gonadotrophins. *The Journal of endocrinology*, **18**(3): 292–304

Evans MJ, Kaufman MH (1981) Establishment in culture of pluripotential cells from mouse embryos. *Nature*, **292**(5819): 154–156

Fong JJ, Nguyen BL, Bridger R, Medrano EE, Wells L, Pan S, Sifers RN (2012) beta-N-Acetylglucosamine (O-GlcNAc) is a novel regulator of mitosis-specific phosphorylations on histone H3. *The Journal of biological chemistry*, **287**(15): 12195–12203

Fonoudi H, Yeganeh M, Fattahi F, Ghazizadeh Z, Rassouli H, Alikhani M, Mojarad BA, Baharvand H, Salekdeh GH, Aghdami N (2013) ISL1 protein transduction promotes cardiomyocyte differentiation from human embryonic stem cells. *PloS one*, **8**(1): e55577

Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, Jensen LJ (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic acids research*, **41**(Database issue): D808–15

Fujiki R, Hashiba W, Sekine H, Yokoyama A, Chikanishi T, Ito S, Imai Y, Kim J, He HH, Igarashi K, Kanno J, Ohtake F, Kitagawa H, Roeder RG, Brown M, Kato S (2011) GlcNAcylation of histone H2B facilitates its monoubiquitination. *Nature*, **480**(7378): 557–560

Gaertner B, Johnston J, Chen K, Wallaschek N, Paulson A, Garruss AS, Gaudenz K, de Kumar B, Krumlauf R, Zeitlinger J (2012) Poised RNA polymerase II changes over developmental time and prepares genes for future expression. *Cell reports*, **2**(6): 1670–1683

Gafni O, Weinberger L, Mansour AA, Manor YS, Chomsky E, Ben-Yosef D, Kalma Y, Viukov S, Maza I, Zviran A, Rais Y, Shipony Z, Mukamel Z, Krupalnik V, Zerbib M, Geula S, Caspi I, Schneir D, Shwartz T, Gilad S, Amann-Zalcenstein D, Benjamin S, Amit I, Tanay A, Massarwa R, Novershtern N, Hanna JH (2013) Derivation of novel human ground state naive pluripotent stem cells. *Nature*, **504**(7479): 282–286

Gambetta MC, Oktaba K, Muller J (2009) Essential role of the glycosyltransferase sxc/Ogt in polycomb repression. *Science (New York, N*, **325**(5936): 93–96

Gao Y, Wells L, Comer FI, Parker GJ, Hart GW (2001) Dynamic O-glycosylation of nuclear and cytosolic proteins: cloning and characterization of a neutral, cytosolic beta-N-acetylglucosaminidase from human brain. *The Journal of biological chemistry*, **276**(13): 9838–9845

Gewinner C, Hart G, Zachara N, Cole R, Beisenherz-Huss C, Groner B (2004) The coactivator of transcription CREB-binding protein interacts preferentially with the glycosylated form of Stat5. *The Journal of biological chemistry*, **279**(5): 3563–3572

Ghosh Z, Wilson KD, Wu Y, Hu S, Quertermous T, Wu JC (2010) Persistent donor cell gene expression among human induced pluripotent stem cells contributes to differences with human embryonic stem cells. *PloS one*, **5**(2): e8975

Gilbert SF (2013) *Developmental Biology*. Sinauer Associates, Inc., 10 edition

Gong CX, Liu F, Iqbal K (2016) O-GlcNAcylation: A regulator of tau pathology and neurodegeneration. *Alzheimer's & dementia : the journal of the Alzheimer's Association*

Guenther MG, Frampton GM, Soldner F, Hockemeyer D, Mitalipova M, Jaenisch R, Young RA (2010) Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. *Cell stem cell*, **7**(2): 249–257

Guo G, von Meyenn F, Santos F, Chen Y, Reik W, Bertone P, Smith A, Nichols J (2016) Naive Pluripotent Stem Cells Derived Directly from Isolated Cells of the Human Inner Cell Mass. *Stem Cell Reports*, **6**(4): 437–446

Guo G, Yang J, Nichols J, Hall JS, Eyres I, Mansfield W, Smith A (2009) Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development (Cambridge, England)*, **136**(7): 1063–1069

Guo JU, Su Y, Zhong C, Ming Gl, Song H (2011) Hydroxylation of 5-Methylcytosine by TET1 Promotes Active DNA Demethylation in the Adult Brain. *Cell*, **145**(3): 423–434. URL `http://linkinghub.elsevier.com/retrieve/pii/S0092867411002996`

Hackett JA, Surani MA (2014) Regulatory principles of pluripotency: from the ground state up. *Cell stem cell*, **15**(4): 416–430

Haltiwanger RS (1998) Modulation of O-Linked N-Acetylglucosamine Levels on Nuclear and Cytoplasmic Proteins in Vivo Using the Peptide O-GlcNAc-beta -N-acetylglucosaminidase Inhibitor O-(2-Acetamido-2-deoxy-Dglucopyranosylidene)amino-N-phenylcarbamate. *Journal of Biological Chemistry*, **273**(6): 3611–3617

Haltiwanger RS, Holt GD, Hart GW (1990) Enzymatic addition of O-GlcNAc to nuclear and cytoplasmic proteins. Identification of a uridine diphospho-N-acetylglucosamine:peptide beta-N-acetylglucosaminyltransferase. *The Journal of biological chemistry*, **265**(5): 2563–2568

Hanna JH, Saha K, Jaenisch R (2010) Pluripotency and Cellular Reprogramming: Facts, Hypotheses, Unresolved Issues. *Cell*, **143**(4): 508–525. URL `http://linkinghub.elsevier.com/retrieve/pii/S009286741001144X`

Hanover JA, Yu S, Lubas WB, Shin SH, Ragano-Caracciola M, Kochran J, Love DC (2003) Mitochondrial and nucleocytoplasmic isoforms of O-linked GlcNAc transferase encoded by a single mammalian gene. *Archives of biochemistry and biophysics*, **409**(2): 287–297

Hart GW, Akimoto Y (2009) The O-GlcNAc modification. In Varki A, Cummings RD, Esko JD (ed.), *Essentials of Glycobiology*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor. URL `http://www.ncbi.nlm.nih.gov/books/NBK1954/`

Hawkins RD, Hon GC, Lee LK, Ngo Q, Lister R, Pelizzola M, Edsall LE, Kuan S, Luu Y, Klugman S, Antosiewicz-Bourget J, Ye Z, Espinoza C, Agarwahl S, Shen L, Ruotti V, Wang W, Stewart R, Thomson JA, Ecker JR, Ren B (2010) Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell stem cell*, **6**(5): 479–491

He Y, Roth C, Turkenburg JP, Davies GJ (2014) Three-dimensional structure of a Streptomyces sviceus GNAT acetyltransferase with similarity to the C-terminal domain of the human GH84 O-GlcNAcase. *Acta crystallographica. Section D, Biological crystallography*, **70**(Pt 1): 186–195

Heckel D, Comtesse N, Brass N, Blin N, Zang KD, Meese E (1998) Novel immunogenic antigen homologous to hyaluronidase in meningioma. *Human molecular genetics*, **7**(12): 1859–1872

Herhaus L (2014) *The regulation of TGF-beta/BMP signalling by deubiquitylating enzymes*. Doctoral Thesis, University of Dundee. URL `http://hdl.handle.net/10588/c53aed5a-920b-4290-a82c-f30b7d807ec6`

Hirami Y, Osakada F, Takahashi K, Okita K, Yamanaka S, Ikeda H, Yoshimura N, Takahashi M (2009) Generation of retinal cells from mouse and human induced pluripotent stem cells. *Neuroscience letters*, **458**(3): 126–131

Hirose Y, Manley JL (2000) RNA polymerase II and the integration of nuclear events. *Genes & development*, **14**(12): 1415–1429

Holland JD, Klaus A, Garratt AN, Birchmeier W (2013) Wnt signaling in stem and cancer stem cells. *Current opinion in cell biology*, **25**(2): 254–264

Huang K, Maruyama T, Fan G (2014) The naive state of human pluripotent stem cells: a synthesis of stem cell and preimplantation embryo transcriptome analyses. *Cell stem cell*, **15**(4): 410–415

Humphrey RK, Beattie GM, Lopez AD, Bucay N, King CC, Firpo MT, Rose-John S, Hayek A (2004) Maintenance of pluripotency in human embryonic stem cells is STAT3 independent. *Stem cells (Dayton, Ohio)*, **22**(4): 522–530

Hunkapiller J, Shen Y, Diaz A, Cagney G, McCleary D, Ramalho-Santos M, Krogan N, Ren B, Song JS, Reiter JF (2012) Polycomb-like 3 promotes polycomb repressive complex 2 binding to CpG islands and embryonic stem cell self-renewal. *PLoS genetics*, **8**(3): e1002576

Hussein SM, Batada NN, Vuoristo S, Ching RW, Autio R, Narva E, Ng S, Sourour M, Hamalainen R, Olsson C, Lundin K, Mikkola M, Trokovic R, Peitz M, Brustle O, Bazett-Jones DP, Alitalo K, Lahesmaa R, Nagy A, Otonkoski T (2011) Copy number variation and selection during reprogramming to pluripotency. *Nature*, **471**(7336): 58–62

Imamura T, Takase M, Nishihara A, Oeda E, Hanai J, Kawabata M, Miyazono K (1997) Smad6 inhibits signalling by the TGF-beta superfamily. *Nature*, **389**(6651): 622–626

Isono T (2011) O-GlcNAc-specific antibody CTD110.6 cross-reacts with N-GlcNAc2-modified proteins induced under glucose deprivation. *PloS one*, **6**(4): e18959

Ito R, Katsura S, Shimada H, Tsuchiya H, Hada M, Okumura T, Sugawara A, Yokoyama A (2013) TET3-OGT interaction increases the stability and the presence of OGT in chromatin. *Genes to cells : devoted to molecular & cellular mechanisms*

Iyer LM, Tahiliani M, Rao A, Aravind L (2009) Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell cycle (Georgetown, Tex*, **8**(11): 1698–1710

James D, Levine AJ, Besser D, Hemmati-Brivanlou A (2005) TGFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. *Development (Cambridge, England)*, **132**(6): 1273–1282

Jang H, Kim TW, Yoon S, Choi SY, Kang TW, Kim SY, Kwon YW, Cho EJ, Youn HD (2012) O-GlcNAc regulates pluripotency and reprogramming by directly acting on core components of the pluripotency network. *Cell stem cell*, **11**(1): 62–74

Jiang J, Chan YS, Loh YH, Cai J, Tong GQ, Lim CA, Robson P, Zhong S, Ng HH (2008) A core Klf circuitry regulates self-renewal of embryonic stem cells. *Nature cell biology*, **10**(3): 353–360

Jinek M, Rehwinkel J, Lazarus BD, Izaurralde E, Hanover JA, Conti E (2004) The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits structural similarities to importin alpha. *Nature structural & molecular biology*, **11**(10): 1001–1007

Kalkan T, Smith A (2014) Mapping the route from naive pluripotency to lineage specification. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **369**(1657)

Kang L, Wang J, Zhang Y, Kou Z, Gao S (2009) iPS cells can support full-term development of tetraploid blastocyst-complemented embryos. *Cell stem cell*, **5**(2): 135–138

Kelly WG, Dahmus ME, Hart GW (1993) RNA polymerase II is a glycoprotein. Modification of the COOH-terminal domain by O-GlcNAc. *The Journal of biological chemistry*, **268**(14): 10416–10424

Khidekel N, Arndt S, Lamarre-Vincent N, Lippert A, Poulin-Kerstien KG, Ramakrishnan B, Qasba PK, Hsieh-Wilson LC (2003) A chemoenzymatic approach toward the rapid and sensitive detection of O-GlcNAc posttranslational modifications. *Journal of the American Chemical Society*, **125**(52): 16162–16163

Khidekel N, Ficarro SB, Clark PM, Bryan MC, Swaney DL, Rexach JE, Sun YE, Coon JJ, Peters EC, Hsieh-Wilson LC (2007) Probing the dynamics of O-GlcNAc glycosylation in the brain using quantitative proteomics. *Nature chemical biology*, **3**(6): 339–348

Khokha MK, Yeh J, Grammer TC, Harland RM (2005) Depletion of three BMP antagonists from Spemann's organizer leads to a catastrophic loss of dorsal structures. *Developmental cell*, **8**(3): 401–411

Kim HS, Park SY, Choi YR, Kang JG, Joo HJ, Moon WK, Cho JW (2009) Excessive O-GlcNAcylation of proteins suppresses spontaneous cardiogenesis in ES cells. *FEBS letters*, **583**(15): 2474–2478

Koch P, Opitz T, Steinbeck J, Ladewig J, Brüstle O (2009) A rosette-type, self-renewing human ES cell-derived neural stem cell with potential for in vitro instruction and synaptic integration. *Proceedings of the National Academy of Sciences of the United States of America*, **106**(9): 3225–3230

Koshida S, Shinya M, Nikaido M, Ueno N, Schulte-Merker S, Kuroiwa A, Takeda H (2002) Inhibition of BMP activity by the FGF signal promotes posterior neural development in zebrafish. *Developmental biology*, **244**(1): 9–20

Kouzarides T (2007) Chromatin modifications and their function. *Cell*, **128**(4): 693–705

Krejci J, Uhlirova R, Galiova G, Kozubek S, Smigova J, Bartova E (2009) Genome-wide reduction in H3K9 acetylation during human embryonic stem cell differentiation. *Journal of cellular physiology*, **219**(3): 677–687

Kreppel LK, Blomberg MA, Hart GW (1997) Dynamic glycosylation of nuclear and cytosolic proteins. Cloning and characterization of a unique O-GlcNAc transferase with multiple tetratricopeptide repeats. *The Journal of biological chemistry*, **272**(14): 9308–9315

Kreppel LK, Hart GW (1999) Regulation of a cytosolic and nuclear O-GlcNAc transferase. Role of the tetratricopeptide repeats. *The Journal of biological chemistry*, **274**(45): 32015–32022

Kumar RM, Cahan P, Shalek AK, Satija R, DaleyKeyser AJ, Li H, Zhang J, Pardee K, Gennert D, Trombetta JJ, Ferrante TC, Regev A, Daley GQ, Collins JJ (2014) Deconstructing transcriptional heterogeneity in pluripotent stem cells. *Nature*, **516**(7529): 56–61

Kunisada Y, Tsubooka-Yamazoe N, Shoji M, Hosoya M (2012) Small molecules induce efficient differentiation into insulin-producing cells from human induced pluripotent stem cells. *Stem Cell Research*, **8**(2): 274–284

Kurek D, Neagu A, Tastemel M, Tuysuz N, Lehmann J, van de Werken, Harmen J G, Philipsen S, van der Linden R, Maas A, van IJcken, Wilfred F J, Drukker M, ten Berge D (2015) Endogenous WNT signals mediate BMP-induced and spontaneous differentiation of epiblast stem cells and human embryonic stem cells. *Stem Cell Reports*, **4**(1): 114–128

Laflamme MA, Chen KY, Naumova AV, Muskheli V, Fugate JA, Dupras SK, Reinecke H, Xu C, Hassanipour M, Police S, O'Sullivan C, Collins L, Chen Y, Minami E, Gill EA, Ueno S, Yuan C, Gold J, Murry CE (2007) Cardiomyocytes derived from human embryonic stem cells in pro-survival factors enhance function of infarcted rat hearts. *Nature biotechnology*, **25**(9): 1015–1024

Lagerlof O, Slocomb JE, Hong I, Aponte Y, Blackshaw S, Hart GW, Huganir RL (2016) The nutrient sensor OGT in PVN neurons regulates feeding. *Science (New York, N.Y.)*, **351**(6279): 1293–1296

Laping NJ, Grygielko E, Mathur A, Butter S, Bomberger J, Tweed C, Martin W, Fornwald J, Lehr R, Harling J, Gaster L, Callahan JF, Olson BA (2002) Inhibition of transforming growth factor (TGF)-beta1-induced extracellular matrix with a novel inhibitor of the TGF-beta type I receptor kinase activity: SB-431542. *Molecular pharmacology*, **62**(1): 58–64

Latham KE, Solter D, Schultz RM (1992) Acquisition of a transcriptionally permissive state during the 1-cell stage of mouse embryogenesis. *Developmental biology*, **149**(2): 457–462

LaVaute TM, Yoo YD, Pankratz MT, Weick JP, Gerstner JR, Zhang SC (2009) Regulation of neural specification from human embryonic stem cells by BMP and FGF. *Stem cells (Dayton, Ohio)*, **27**(8): 1741–1749

Lawson KA, Dunn NR, Roelen BA, Zeinstra LM, Davis AM, Wright CV, Korving JP, Hogan BL (1999) Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes & Development*, **13**(4): 424–436

Lazarus MB, Nam Y, Jiang J, Sliz P, Walker S (2011) Structure of human O-GlcNAc transferase and its complex with a peptide substrate. *Nature*, **469**(7331): 564–567

Leeb M, Wutz A (2007) Ring1B is crucial for the regulation of developmental control genes and PRC1 proteins but not X inactivation in embryonic cells. *The Journal of cell biology*, **178**(2): 219–229

Lister R, Pelizzola M, Kida YS, Hawkins RD, Nery JR, Hon G, Antosiewicz-Bourget J, O'Malley R, Castanon R, Klugman S, Downes M, Yu R, Stewart R, Ren B, Thomson JA, Evans RM, Ecker JR (2011) Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature*, **471**(7336): 68–73

Love DC, Kochan J, Cathey RL, Shin SH, Hanover JA (2003) Mitochondrial and nucleocytoplasmic targeting of O-linked GlcNAc transferase. *Journal of cell science*, **116**(Pt 4): 647–654

Lubas WA, Frank DW, Krause M, Hanover JA (1997) O-Linked GlcNAc transferase is a conserved nucleocytoplasmic protein containing tetratricopeptide repeats. *The Journal of biological chemistry*, **272**(14): 9316–9324

Lubas WA, Hanover JA (2000) Functional expression of O-linked GlcNAc transferase. Domain structure and substrate specificity. *The Journal of biological chemistry*, **275**(15): 10983–10988

MacCord K (2013) Germ Layers. URL `http://embryo.asu.edu/handle/10776/6273`

Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, Firth A, Singer O, Trono D, Pfaff SL (2012) Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature*, **487**(7405): 57–63

Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S (2002) The protein kinase complement of the human genome. *Science (New York, N*, **298**(5600): 1912–1934

Marchetto MCN, Yeo GW, Kainohana O, Marsala M, Gage FH, Muotri AR (2009) Transcriptional signature and memory retention of human-induced pluripotent stem cells. *PloS one*, **4**(9): e7076

Marks H, Kalkan T, Menafra R, Denissov S, Jones K, Hofemeister H, Nichols J, Kranz A, Stewart AF, Smith A, Stunnenberg HG (2012) The transcriptional and epigenomic foundations of ground state pluripotency. *Cell*, **149**(3): 590–604

Martin GR (1981) Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, **78**(12): 7634–7638

Martin GR, Wiley LM, Damjanov I (1977) The development of cystic embryoid bodies in vitro from clonal teratocarcinoma stem cells. *Developmental biology*, **61**(2): 230–244

Martin Gonzalez J, Morgani SM, Bone RA, Bonderup K, Abelchian S, Brakebusch C, Brickman JM (2016) Embryonic Stem Cell Culture Conditions Support Distinct States Associated with Different Developmental Stages and Potency. *Stem Cell Reports*, **7**(2): 177–191

Matsuda T, Nakamura T, Nakao K, Arai T, Katsuki M, Heike T, Yokota T (1999) STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. *The EMBO journal*, **18**(15): 4261–4269

Maury JJP, El Farran CA, Ng D, Loh YH, Bi X, Bardor M, Choo ABH (2015) RING1B O-GlcNAcylation regulates gene targeting of polycomb repressive complex 1 in human embryonic stem cells. *Stem Cell Research*, **15**(1): 182–189

Maury JJP, Ng D, Bi X, Bardor M, Choo ABH (2013) MULTIPLE REACTION MONITORING MASS SPECTROMETRY FOR THE DISCOVERY AND QUANTIFICATION OF O-GLCNAC MODIFIED PROTEINS. *Analytical chemistry*

Medine CN, Lucendo-Villarin B, Zhou W, West CC, Hay DC (2011) Robust Generation of Hepatocyte-like Cells from Human Embryonic Stem Cell Populations. *Journal of Visualized Experiments*, (56)

Morgani SM, Canham MA, Nichols J, Sharov AA, Migueles RP, Ko MSH, Brickman JM (2013) Totipotent embryonic stem cells arise in ground-state culture conditions. *Cell reports*, **3**(6): 1945–1957

Myers SA, Panning B, Burlingame AL (2011) Polycomb repressive complex 2 is necessary for the normal site-specific O-GlcNAc distribution in mouse embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, **108**(23): 9490–9495

Nakagawa M, Koyanagi M, Tanabe K, Takahashi K, Ichisaka T, Aoi T, Okita K, Mochiduki Y, Takizawa N, Yamanaka S (2008) Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nature biotechnology*, **26**(1): 101–106

Nakao A, Afrakhte M, Moren A, Nakayama T, Christian JL, Heuchel R, Itoh S, Kawabata M, Heldin NE, Heldin CH, ten Dijke P (1997) Identification of Smad7, a TGFbeta-inducible antagonist of TGF-beta signalling. *Nature*, **389**(6651): 631–635

Neely MD, Litt MJ, Tidball AM, Li GG, Aboud AA, Hopkins CR, Chamberlin R, Hong CC, Ess KC, Bowman AB (2012) DMH1, a highly selective small molecule BMP inhibitor promotes neurogenesis of hiPSCs: comparison of PAX6 and SOX1 expression during neural induction. *ACS chemical neuroscience*, **3**(6): 482–491

Niwa H, Burdon T, Chambers I, Smith A (1998) Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes & development*, **12**(13): 2048–2060

Niwa H, Masui S, Chambers I, Smith AG, Miyazaki Ji (2002) Phenotypic complementation establishes requirements for specific POU domain and generic transactivation function of Oct-3/4 in embryonic stem cells. *Molecular and cellular biology*, **22**(5): 1526–1536

Okita K, Yamanaka S (2006) Intracellular signaling pathways regulating pluripotency of embryonic stem cells. *Current stem cell research & therapy*, **1**(1): 103–111

Olivier-Van Stichelen S, Guinez C, Mir AM, Perez-Cervera Y, Liu C, Michalski JC, Lefebvre T (2012) The hexosamine biosynthetic pathway and O-GlcNAcylation drive the expression of beta-catenin and cell proliferation. *American journal of physiology*, **302**(4): E417–24

Oshimori N, Fuchs E (2012) The harmonies played by TGF-beta in stem cell biology. *Cell stem cell*, **11**(6): 751–764

Papapetrou EP, Tomishima MJ, Chambers SM, Mica Y, Reed E, Menon J, Tabar V, Mo Q, Studer L, Sadelain M (2009) Stoichiometric and temporal requirements of Oct4, Sox2, Klf4, and c-Myc expression for efficient human iPSC induction and differentiation. *Proceedings of the National Academy of Sciences of the United States of America*, **106**(31): 12759–12764

Pastor WA, Di Chen, Liu W, Kim R, Sahakyan A, Lukianchikov A, Plath K, Jacobsen SE, Clark AT (2016) Naive Human Pluripotent Cells Feature a Methylation Landscape Devoid of Blastocyst or Germline Memory. *Cell stem cell*, **18**(3): 323–329

Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic acids research*, **29**(9): e45

Proffitt KD, Madan B, Ke Z, Pendharkar V, Ding L, Lee MA, Hannoush RN, Virshup DM (2013) Pharmacological inhibition of the Wnt acyltransferase PORCN prevents growth of WNT-driven mammary cancer. *Cancer research*, **73**(2): 502–507

Qi X, Li TG, Hao J, Hu J, Wang J, Simmons H, Miura S, Mishina Y, Zhao GQ (2004) BMP4 supports self-renewal of embryonic stem cells by inhibiting mitogen-activated protein kinase pathways. *Proceedings of the National Academy of Sciences of the United States of America*, **101**(16): 6027–6032

Qiu D, Ye S, Ruiz B, Zhou X, Liu D, Zhang Q, Ying QL (2015) Klf2 and Tfcp2l1, Two Wnt/beta-Catenin Targets, Act Synergistically to Induce and Maintain Naive Pluripotency. *Stem Cell Reports*, **5**(3): 314–322

Rao FV, Schuttelkopf AW, Dorfmueller HC, Ferenbach AT, Navratilova I, van Aalten, Daan M F (2013) Structure of a bacterial putative acetyltransferase defines the fold of the human O-GlcNAcase C-terminal domain. *Open Biology*, **3**(10): 130021

Raz R, Lee CK, Cannizzaro LA, d'Eustachio P, Levy DE (1999) Essential role of STAT3 for embryonic stem cell pluripotency. *Proceedings of the National Academy of Sciences of the United States of America*, **96**(6): 2846–2851

Rexach JE, Rogers CJ, Yu SH, Tao J, Sun YE, Hsieh-Wilson LC (2010) Quantification of O-glycosylation stoichiometry and dynamics using resolvable mass tags. *Nature chemical biology*, **6**(9): 645–651

Robertson EJ (1986) Pluripotential stem cell lines as a route into the mouse germ line. *Trends in Genetics*, **2**: 9–13

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*, **26**(1): 139–140

Sacco F, Perfetto L, Castagnoli L, Cesareni G (2012) The human phosphatase interactome: An intricate family portrait. *FEBS letters*, **586**(17): 2732–2739

Sakabe K, Wang Z, Hart GW (2010) Beta-N-acetylglucosamine (O-GlcNAc) is part of the histone code. *Proceedings of the National Academy of Sciences of the United States of America*, **107**(46): 19915–19920

Sato N, Meijer L, Skaltsounis L, Greengard P, Brivanlou AH (2004) Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. *Nature medicine*, **10**(1): 55–63

Sayat R, Leber B, Grubac V, Wiltshire L, Persad S (2008) O-GlcNAc-glycosylation of beta-catenin regulates its nuclear localization and transcriptional activity. *Experimental cell research*, **314**(15): 2774–2787

Schultz RM (1993) Regulation of zygotic gene activation in the mouse. *BioEssays : news and reviews in molecular, cellular and developmental biology*, **15**(8): 531–538

Schultz RM (2002) The molecular foundations of the maternal to zygotic transition in the preimplantation embryo. *Human Reproduction Update*, **8**(4): 323–331

Shafi R, Iyer SP, Ellies LG, O'Donnell N, Marek KW, Chui D, Hart GW, Marth JD (2000) The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny. *Proceedings of the National Academy of Sciences of the United States of America*, **97**(11): 5735–5739

Shen X, Liu Y, Hsu YJ, Fujiwara Y, Kim J, Mao X, Yuan GC, Orkin SH (2008) EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Molecular cell*, **32**(4): 491–502

Shi FT, Kim H, Lu W, He Q, Liu D, Goodell MA, Wan M, Songyang Z (2013) Ten-eleven translocation 1 (Tet1) is regulated by O-linked N-acetylglucosamine transferase (Ogt) for target gene repression in mouse embryonic stem cells. *The Journal of biological chemistry*, **288**(29): 20776–20784

Silva J, Barrandon O, Nichols J, Kawaguchi J, Theunissen TW, Smith A (2008) Promotion of reprogramming to ground state pluripotency by signal inhibition. *PLoS biology*, **6**(10): e253

Sinclair DAR, Syrzycka M, Macauley MS, Rastgardani T, Komljenovic I, Vocadlo DJ, Brock HW, Honda BM (2009) Drosophila O-GlcNAc transferase (OGT) is encoded by the Polycomb group (PcG) gene, super sex combs (sxc). *Proceedings of the National Academy of Sciences of the United States of America*, **106**(32): 13427–13432

Smith AG, Heath JK, Donaldson DD, Wong GG, Moreau J, Stahl M, Rogers D (1988) Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature*, **336**(6200): 688–690

Smith JR, Vallier L, Lupo G, Alexander M, Harris WA, Pedersen RA (2008) Inhibition of Activin/Nodal signaling promotes specification of human embryonic stem cells into neuroectoderm. *Developmental biology*, **313**(1): 107–117

Snow CM (1987) Monoclonal antibodies identify a group of nuclear pore complex glycoproteins. *The Journal of cell biology*, **104**(5): 1143–1156

Speakman CM, Domke TC, Wongpaiboonwattana W, Sanders K, Mudaliar M, van Aalten, Daan M F, Barton GJ, Stavridis MP (2014) Elevated O-GlcNAc levels activate epigenetically repressed genes and delay mouse ESC differentiation without affecting naive to primed cell transition. *Stem cells (Dayton, Ohio)*, **32**(10): 2605–2615

Stevens LC, Little CC (1954) Spontaneous Testicular Teratomas in an Inbred Strain of Mice. *Proceedings of the National Academy of Sciences of the United States of America*, **40**(11): 1080–1087

Sumi T, Fujimoto Y, Nakatsuji N, Suemori H (2004) STAT3 is dispensable for maintenance of self-renewal in nonhuman primate embryonic stem cells. *Stem cells (Dayton, Ohio)*, **22**(5): 861–872

Supek F, Bosnjak M, Skunca N, Smuc T (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS one*, **6**(7): e21800

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A (2009) Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science*, **324**(5929): 930–935

Tai HC, Khidekel N, Ficarro SB, Peters EC, Hsieh-Wilson LC (2004) Parallel identification of O-GlcNAc-modified proteins from cell lysates. *Journal of the American Chemical Society*, **126**(34): 10500–10501

Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, **131**(5): 861–872

Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, **126**(4): 663–676

Takahashi K, Yamanaka S (2016) A decade of transcription factor-mediated reprogramming to pluripotency. *Nature reviews. Molecular cell biology*, **17**(3): 183–193

Takashima Y, Guo G, Loos R, Nichols J, Ficz G, Krueger F, Oxley D, Santos F, Clarke J, Mansfield W, Reik W, Bertone P, Smith A (2014) Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell*, **158**(6): 1254–1269

Tanaka T, Tohyama S, Murata M, Nomura F, Kaneko T, Chen H, Hattori F, Egashira T, Seki T, Ohno Y, Koshimizu U, Yuasa S, Ogawa S, Yamanaka S, Yasuda K, Fukuda K (2009) In vitro pharmacologic testing using human induced pluripotent stem cell-derived cardiomyocytes. *Biochemical and biophysical research communications*, **385**(4): 497–502

Taura D, Noguchi M, Sone M, Hosoda K, Mori E, Okada Y, Takahashi K, Homma K, Oyamada N, Inuzuka M, Sonoyama T, Ebihara K, Tamura N, Itoh H, Suemori H, Nakatsuji N, Okano H, Yamanaka S, Nakao K (2009) Adipogenic differentiation of human induced pluripotent stem cells: comparison with that of human embryonic stem cells. *FEBS letters*, **583**(6): 1029–1033

Teo AKK, Ali Y, Wong KY, Chipperfield H, Sadasivam A, Poobalan Y, Tan EK, Wang ST, Abraham S, Tsuneyoshi N, Stanton LW, Dunn NR (2012) Activin and BMP4 synergistically promote formation of definitive endoderm in human embryonic stem cells. *Stem cells (Dayton, Ohio)*, **30**(4): 631–642

Tesar PJ, Chenoweth JG, Brook FA, Davies TJ, Evans EP, Mack DL, Gardner RL, McKay RDG (2007) New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature*, **448**(7150): 196–199

Theunissen TW, Powell BE, Wang H, Mitalipova M, Faddah DA, Reddy J, Fan ZP, Maetzel D, Ganz K, Shi L, Lungjangwa T, Imsoonthornruksa S, Stelzer Y, Rangarajan S, D'Alessio A, Zhang J, Gao Q, Dawlaty MM, Young RA, Gray NS, Jaenisch R (2014) Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. *Cell stem cell*, **15**(4): 471–487

Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM (1998) Embryonic stem cell lines derived from human blastocysts. *Science (New York, N*, **282**(5391): 1145–1147

Timmer JR, Wang C, Niswander L (2002) BMP signaling patterns the dorsal and intermediate neural tube via regulation of homeobox and helix-loop-helix transcription factors. *Development*, **129**(10): 2459

Tokuzawa Y, Kaiho E, Maruyama M, Takahashi K, Mitsui K, Maeda M, Niwa H, Yamanaka S (2003) Fbx15 is a novel target of Oct3/4 but is dispensable for embryonic stem cell self-renewal and mouse development. *Molecular and cellular biology*, **23**(8): 2699–2708

Toleman C, Paterson AJ, Whisenhunt TR, Kudlow JE (2004) Characterization of the histone acetyltransferase (HAT) domain of a bifunctional protein with activable O-GlcNAcase and HAT activities. *The Journal of biological chemistry*, **279**(51): 53665–53673

Torres CR, Hart GW (1984) Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc. *The Journal of biological chemistry*, **259**(5): 3308–3317

Vallier L, Touboul T, Chng Z, Brimpari M, Hannan N, Millan E, Smithers LE, Trotter M, Rugg-Gunn P, Weber A, Pedersen RA (2009) Early cell fate decisions of human embryonic stem cells and mouse epiblast stem cells are controlled by the same signalling pathways. *PloS one*, **4**(6): e6082

Vella P, Scelfo A, Jammula S, Chiacchiera F, Williams K, Cuomo A, Roberto A, Christensen J, Bonaldi T, Helin K, Pasini D (2013) Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. *Molecular cell*, **49**(4): 645–656

Vocadlo DJ, Hang HC, Kim EJ, Hanover JA, Bertozzi CR (2003) A chemical approach for identifying O-GlcNAc-modified proteins in cells. *Proceedings of the National Academy of Sciences of the United States of America*, **100**(16): 9116–9121

Walker E, Chang WY, Hunkapiller J, Cagney G, Garcha K, Torchia J, Krogan NJ, Reiter JF, Stanford WL (2010) Polycomb-like 2 associates with PRC2 and regulates transcriptional networks during mouse embryonic stem cell self-renewal and differentiation. *Cell stem cell*, **6**(2): 153–166

Wang G, Zhang H, Zhao Y, Li J, Cai J, Wang P, Meng S, Feng J, Miao C, Ding M, Li D, Deng H (2005) Noggin and bFGF cooperate to maintain the pluripotency of human embryonic stem cells in the absence of feeder layers. *Biochemical and biophysical research communications*, **330**(3): 934–942

Ware CB, Nelson AM, Mecham B, Hesson J, Zhou W, Jonlin EC, Jimenez-Caliani AJ, Deng X, Cavanaugh C, Cook S, Tesar PJ, Okada J, Margaretha L, Sperber H, Choi M, Blau CA, Treuting PM, Hawkins RD, Cirulli V, Ruohola-Baker H (2014) Derivation of naive human embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, **111**(12): 4484–4489

Warrier S, Popovic M, van der Jeught M, Heindryckx B (2016) Establishment and Characterization of Naive Pluripotency in Human Embryonic Stem Cells. *Methods in molecular biology (Clifton, N.J.)*

Wells L, Vosseller K, Hart GW (2003) A role for N-acetylglucosamine as a nutrient sensor and mediator of insulin resistance. *Cellular and molecular life sciences : CMLS*, **60**(2): 222–228

Wen B, Wu H, Shinkai Y, Irizarry RA, Feinberg AP (2009) Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. *Nature genetics*, **41**(2): 246–250

Wernig M, Meissner A, Foreman R, Brambrink T, Ku M, Hochedlinger K, Bernstein BE, Jaenisch R (2007) In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature*, **448**(7151): 318–324

Whelan SA, Lane MD, Hart GW (2008) Regulation of the O-linked beta-N-acetylglucosamine transferase by insulin signaling. *The Journal of biological chemistry*, **283**(31): 21411–21417

Whisenhunt TR, Yang X, Bowe DB, Paterson AJ, van Tine BA, Kudlow JE (2006) Disrupting the enzyme complex regulating O-GlcNAcylation blocks signaling and development. *Glycobiology*, **16**(6): 551–563

Williams RL, Hilton DJ, Pease S, Willson TA, Stewart CL, Gearing DP, Wagner EF, Metcalf D, Nicola NA, Gough NM (1988) Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature*, **336**(6200): 684–687

Wilson PA, Lagna G, Suzuki A, Hemmati-Brivanlou A (1997) Concentration-dependent patterning of the Xenopus ectoderm by BMP4 and its signal transducer Smad1. *Development (Cambridge, England)*, **124**(16): 3177–3184

Wu H, Zhang Y (2011) Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Genes & development*, **25**(23): 2436–2452

Wu MY, Hill CS (2009) Tgf-beta superfamily signaling in embryonic development and homeostasis. *Developmental cell*, **16**(3): 329–343

Xu RH, Chen X, Li DS, Li R, Addicks GC, Glennon C, Zwaka TP, Thomson JA (2002) BMP4 initiates human embryonic stem cell differentiation to trophoblast. *Nature biotechnology*, **20**(12): 1261–1264

Xu RH, Peck RM, Li DS, Feng X, Ludwig T, Thomson JA (2005) Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells. *Nature methods*, **2**(3): 185–190

Xu RH, Sampsell-Barron TL, Gu F, Root S, Peck RM, Pan G, Yu J, Antosiewicz-Bourget J, Tian S, Stewart R, Thomson JA (2008) NANOG is a direct target of TGFbeta/activin-mediated SMAD signaling in human ESCs. *Cell stem cell*, **3**(2): 196–206

Yang X, Zhang F, Kudlow JE (2002) Recruitment of O-GlcNAc Transferase to Promoters by Corepressor mSin3A. *Cell*, **110**(1): 69–80

Ying QL, Nichols J, Chambers I, Smith A (2003a) BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell*, **115**(3): 281–292

Ying QL, Stavridis M, Griffiths D, Li M, Smith A (2003b) Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nature biotechnology*, **21**(2): 183–186

Ying QL, Wray J, Nichols J, Batlle-Morera L, Doble B, Woodgett J, Cohen P, Smith A (2008) The ground state of embryonic stem cell self-renewal. *Nature*, **453**(7194): 519–523

Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, Nie J, Jonsdottir GA, Ruotti V, Stewart R, Slukvin II, Thomson JA (2007) Induced pluripotent stem cell lines derived from human somatic cells. *Science (New York, N.Y.)*, **318**(5858): 1917–1920

Yuzwa SA, Macauley MS, Heinonen JE, Shan X, Dennis RJ, He Y, Whitworth GE, Stubbs KA, McEachern EJ, Davies GJ, Vocadlo DJ (2008) A potent mechanism-inspired O-GlcNAcase inhibitor that blocks phosphorylation of tau in vivo. *Nature chemical biology*, **4**(8): 483–490

Zhang J, Klos M, Wilson GF, Herman AM, Lian X, Raval KK, Barron MR, Hou L, Soerens AG, Yu J, Palecek SP, Lyons GE, Thomson JA, Herron TJ, Jalife J, Kamp TJ (2012) Extracellular matrix promotes highly efficient cardiac differentiation of human pluripotent stem cells: the matrix sandwich method. *Circulation research*, **111**(9): 1125–1136

Zhang Q, Liu X, Gao W, Li P, Hou J, Li J, Wong J (2014) Differential regulation of the ten-eleven translocation (TET) family of dioxygenases by O-linked beta-N-acetylglucosamine transferase (OGT). *The Journal of biological chemistry*, **289**(9): 5986–5996

Zhang S, Roche K, Nasheuer HP, Lowndes NF (2011) Modification of Histones by Sugar -N-Acetylglucosamine (GlcNAc) Occurs on Multiple Residues, Including Histone H3 Serine 10, and Is Cell Cycle-regulated. *Journal of Biological Chemistry*, **286**(43): 37483–37495

Zhao Xy, Li W, Lv Z, Liu L, Tong M, Hai T, Hao J, Guo Cl, Ma Qw, Wang L, Zeng F, Zhou Q (2009) iPS cells produce viable mice through tetraploid complementation. *Nature*, **461**(7260): 86–90